

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:42 ; Search time 48.0351 Seconds
(without alignments)
3958.655 Million cell updates/sec

Title: US-09-989-981A-8
Perfect score: 3506
Sequence: 1 MAGKAAEERGLPKGATPQDT.....FMVLYYVSLRFIKQKPSQDW 673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	3506	100.0	673	6	AAE31705	Aae31705 Human ABC
2	3502	99.9	673	5	ABP52129	Abp52129 Homo sapi
3	2888.5	82.4	672	6	AAE31703	Aae31703 Mouse ABC
4	1961	55.9	374	5	ABG61539	Abg61539 Human tra
5	730.5	20.8	632	3	AAG18079	Aag18079 Arabidops
6	730.5	20.8	648	3	AAG18078	Aag18078 Arabidops
7	724	20.7	625	3	AAG18080	Aag18080 Arabidops
8	713	20.3	652	5	AAU96986	Aau96986 Rat ABCG5
9	705	20.1	651	5	AAU96990	Aau96990 Human ABC

10	697	19.9	651	5	AAU96993	Aau96993	Human	ABC
11	697	19.9	651	5	AAU96984	Aau96984	Human	ABC
12	697	19.9	651	5	AAE13290	Aae13290	Human	sit
13	697	19.9	651	6	AAE31704	Aae31704	Human	ABC
14	696	19.9	651	5	AAU96989	Aau96989	Human	ABC
15	694	19.8	651	5	AAU96992	Aau96992	Human	ABC
16	691.5	19.7	652	5	AAU96985	Aau96985	Mouse	ABC
17	688.5	19.6	652	5	AAE13308	Aae13308	Mouse	sit
18	688.5	19.6	652	5	AAE13309	Aae13309	Mouse	sit
19	688.5	19.6	652	5	AAE13289	Aae13289	Mouse	sit
20	688.5	19.6	652	6	AAE31702	Aae31702	Mouse	ABC
21	675	19.3	649	5	ABP52128	Abp52128	Homo sapi	
22	666	19.0	657	5	ABB07272	Abb07272	Murine	BC
23	665	19.0	687	3	AAY78981	Aay78981	Silkworm	
24	656	18.7	687	4	ABB59384	Abb59384	Drosophil	
25	642.5	18.3	655	4	AAB60104	Aab60104	Human	tra
26	642.5	18.3	655	5	AAO14781	Aao14781	Human	BCR
27	642.5	18.3	655	5	AAU80028	Aau80028	Human	ABC
28	642.5	18.3	655	6	ABR58077	Abr58077	Human	ABC
29	642.5	18.3	655	6	ADA10917	Ada10917	Human	cDN
30	642.5	18.3	655	7	ADC54182	Adc54182	Human	bre
31	640.5	18.3	655	5	AAU80029	Aau80029	Human	ABC
32	640.5	18.3	663	2	AAU15221	Aay15221	Breast	Ca
33	638.5	18.2	655	3	AAY95365	Aay95365	ATP-bindi	
34	638.5	18.2	655	4	AAU04348	Aau04348	Human	BCR
35	638.5	18.2	655	5	ABP52127	Abp52127	Homo sapi	
36	638.5	18.2	655	5	ABB07270	Abb07270	Human	BCR
37	638.5	18.2	655	6	ABU63376	Abu63376	Human	mit
38	638.5	18.2	665	5	AAO14782	Aao14782	Human	BCR
39	638.5	18.2	665	5	AAO14783	Aao14783	Human	BCR
40	634.5	18.1	655	5	ABB07273	Abb07273	Human	BCR
41	621	17.7	666	5	ABB57112	Abb57112	Mouse	isc
42	620	17.7	662	6	ABO07271	Abo07271	Human	p53
43	620	17.7	663	6	ABB82647	Abb82647	Human	Dev
44	618	17.6	638	5	ABB98349	Abb98349	Human	ABC
45	617	17.6	674	5	ABP52126	Abp52126	Homo sapi	

ALIGNMENTS

RESULT 1

AAE31705

ID AAE31705 standard; protein; 673 AA.

XX

AC AAE31705;

XX

DT 24-MAR-2003 (first entry)

XX

DE Human ABCG8 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;

KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;

KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;

KW ABCG5.

XX

OS Homo sapiens.
 XX
 PN WO200281691-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 20-NOV-2001; 2001WO-US043823.
 XX
 PR 20-NOV-2000; 2000US-0252235P.
 PR 28-NOV-2000; 2000US-0253645P.
 XX
 PA (TULA-) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 XX
 DR WPI; 2003-058548/05.
 DR N-PSDB; AAD48883.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 22; Page 81-82; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
 CC are useful for treating or preventing sterol-related disorders such as
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also
 CC useful in gene therapy. The present sequence is human ABCG8 protein
 XX
 SQ Sequence 673 AA;

Query Match 100.0%; Score 3506; DB 6; Length 673;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAGKAAEERGLPKGATPQDTSGLQDRLFSSSESDNSLYFTYSGQPNTLEVRDLNYQVDLAS	60
Db	1	MAGKAAEERGLPKGATPQDTSGLQDRLFSSSESDNSLYFTYSGQPNTLEVRDLNYQVDLAS	60
Qy	61	QVPWFQQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITG	120
Db	61	QVPWFQQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITG	120
Qy	121	RGHGGKIKSGQIWINQPPSSPQLVRKCVAVHRQHNQLLPNLTVRETAFIAQMRLPRTFS	180
Db	121	RGHGGKIKSGQIWINQPPSSPQLVRKCVAVHRQHNQLLPNLTVRETAFIAQMRLPRTFS	180
Qy	181	QAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPT	240
Db	181	QAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPT	240

Qy	241	SGLDSFTAHLNVLKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM	300
Db	241	SGLDSFTAHLNVLKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM	300
Qy	301	VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL	360
Db	301	VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL	360
Qy	361	WKAETKDLDEDTCEVSSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH	420
Db	361	WKAETKDLDEDTCEVSSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH	420
Qy	421	GAEACLSMTIGFLYFGHSGSIQLSFMDTAALLFMIGALIPFNVILDVISKYSERAMLYY	480
Db	421	GAEACLSMTIGFLYFGHSGSIQLSFMDTAALLFMIGALIPFNVILDVISKYSERAMLYY	480
Qy	481	ELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWL VVF	540
Db	481	ELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWL VVF	540
Qy	541	CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF	600
Db	541	CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF	600
Qy	601	EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV	660
Db	601	EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV	660
Qy	661	SLRFIKQKPSQDW	673
Db	661	SLRFIKQKPSQDW	673

RESULT 2

ABP52129

ID ABP52129 standard; protein; 673 AA.

XX

AC ABP52129;

XX

DT 10-OCT-2002 (first entry)

XX

DE Homo sapiens ABC transporter ABCG8 protein SEQ ID NO:81.

XX

KW ATP-binding cassette transporter; ABC transporter; modulation; D loop;

KW cancer; bacterial infection; fungal infection; protozoal infection;

KW antibacterial; fungicide; protozoacide.

XX

OS Homo sapiens.

XX

PN EP1217066-A1.

XX

PD 26-JUN-2002.

XX

PF 21-DEC-2000; 2000EP-00870316.

XX

PR 21-DEC-2000; 2000EP-00870316.

XX

PA (UYGE-) UNIV GENT.

XX

DR WPI; 2002-550404/59.

XX

PT Modulating activity of ATP-binding cassette (ABC) transporters by
PT influencing dimerization of nucleotide binding domains through use of D
PT loop sequence of an ABC transporter, or its antisense peptide or peptide
PT mimetic.

XX

PS Disclosure; Fig 3; 290pp; English.

XX

CC The present invention describes a method (M1) for modulating the activity
CC of ATP-binding cassette (ABC) transporters by influencing the
CC dimerisation of the nucleotide binding domains comprises using: (a) a
CC polypeptide (polyP) consisting of 5-50 amino acids comprising the D loop
CC sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyP
CC consisting of the D loop sequence of an ABC transporter; (c) a peptide
CC mimetic or antisense peptide of (a) or (b). ABC transporters have
CC antibacterial, fungicide and protozoacide activities. (M1) is useful for
CC selectively modulating the activity of ABC transporters belonging to the
CC group of multidrug transporter/P-glycoproteins. Bacterial, fungal or
CC protozoal ABC transporters are involved in the infection of a mammal or
CC in the induction of resistance to antibiotics or drugs in a mammal. (M1)
CC is useful for preventing, treating or alleviating diseases associated
CC with functionality of an ABC transporter. ABP52092 to ABP52140 represent
CC ABC transporter proteins given in the exemplification of the present
CC invention

XX

SQ Sequence 673 AA;

Query Match 99.9%; Score 3502; DB 5; Length 673;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLAS	60
Db	1	MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLAS	60
Qy	61	QVPWFQQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVTIG	120
Db	61	QVPWFQQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVTIG	120
Qy	121	RGHGGKIKSGQIWINQPSQPQLVRKCAHVVRQHNQLLPNLTVRETAFIAQMRLPRTFS	180
Db	121	RGHGGKIKSGQIWINQPSQPQLVRKCAHVVRQHNQLLPNLTVRETAFIAQMRLPRTFS	180
Qy	181	QAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPT	240
Db	181	QAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPT	240
Qy	241	SGLDSFTAHLNVLKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGA AQHM	300
Db	241	SGLDSFTAHLNVLKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGA AQHM	300
Qy	301	VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL	360
Db	301	VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL	360

Qy 361 WKAETKDLDEDTCEVSSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH 420
 |||
 Db 361 WKAETKDLDEDTCEVSSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH 420

Qy 421 GAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY 480
 |||
 Db 421 GAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY 480

Qy 481 ELEDGLYTTGPTYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWLTVV 540
 |||
 Db 481 ELEDGLYTTGPTYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWLTVV 540

Qy 541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWCF 600
 |||
 Db 541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWCF 600

Qy 601 EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV 660
 |||
 Db 601 EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV 660

Qy 661 SLRFIKQKPSQDW 673
 |||
 Db 661 SLRFIKQKPSQDW 673

RESULT 3

AAE31703

ID AAE31703 standard; protein; 672 AA.

XX

AC AAE31703;

XX

DT 24-MAR-2003 (first entry)

XX

DE Mouse ABCG8 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
 KW ABCG5.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT Misc-difference 440

FT /note= "Encoded by AAG"

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.

XX

PF 20-NOV-2001; 2001WO-US043823.

XX

PR 20-NOV-2000; 2000US-0252235P.

PR 28-NOV-2000; 2000US-0253645P.

XX

Db 361 WKAEAKELNTSTHTVSLTLTQDTDC-GTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIH 419

Qy	421	GAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVI	LDV	ISKY	SERAMLYY	480
		: : : :				
Db	420	GSEACLMSLIIGFLYYGHGALQLSFMDTAALLFMIGALIPFNVI	LDV	VSKCH	SERSMLYY	479
Qy	481	ELEDGLYTTGPYFFAKILGELPEHCAYII	IYG	MPTYWLANLRPGLQPFLLHFLLVWLVVF		540
					: :	
Db	480	ELEDGLYTAGPYFFAKILGELPEHCAYVII	IYAMPIYWL	TNLRPVP	ELFLLHFLLVWLVVF	539
Qy	541	CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSS	SLW	TPAWISKVSFLRWCF		600
		: : :			: :	
Db	540	CCRTMALAASAMLPTFHMSFFCNALYNSFYLTAGFMINLDNL	WIVPAWISKLSFLRWCF			599
Qy	601	EGLMKIQFSRRTYKMPLGNLTI	AVSGDKILS	AMELDSYPLYAIYLIVIGLSGGFMVLYYV		660
		: : : :		: : : :	: : :	
Db	600	SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLNSHPLYAIYLIVIGISYGF	FLY	YL		659
Qy	661	SLRFIKQKPSQDW				673
		:				
Db	660	SLKLIKOKSIQDW				672

ABG61539

ID ABG61539 standard; protein; 374 AA.

XX

AC ABG61539;

XX

DT 27-AUG-2002 (first entry)

XX

DE Human transporter and ion channel, TRICH9, Incyte ID 6585710CD1.

XX

KW Human; transporter and ion channel; TRICH; transport disorder;
KW neurological disorder; muscle disorder; immunological disorder; cancer;
KW scleroderma; systemic lupus erythematosus; allergy; leukaemia;
KW cell proliferative disorder; cervical cancer; breast cancer;
KW neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
KW myotonic dystrophy; catatonia; endocrine disorder; diabetes;
KW Grave's disease; gastrointestinal disorder; Crohn's disease;
KW renal disorder; Good pasture's syndrome; viral infection; cirrhosis;
KW bacterial infection; fungal infection; parasitic infection;
KW protozoal infection; helminthic infection; cardiovascular disorder;
KW atherosclerosis; hepatic disease.

XX

OS Homo sapiens.

XX

PN WO200240541-A2.

XX

PD 23-MAY-2002.

XX

PF 25-OCT-2001; 2001WO-US046055.

XX

PR 27-OCT-2000; 2000US-0243989P.

PR 03-NOV-2000; 2000US-0245904P.

PR 09-NOV-2000; 2000US-0247673P.

PR 17-NOV-2000; 2000US-0249661P.

PR 20-NOV-2000; 2000US-0252232P.

PR 01-DEC-2000; 2000US-0250790P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Tang YT, Yue H, Nguyen DB, Hafalia AJA, Elliott VS, Lu Y;
PI Walia NK, Yao MG, Baughn MR, Gandhi AR, Ding L, Sanjanwala M;
PI Ramkumar J, Arvizu C, Gietzen KJ, Lal PG, Azimzai Y, Khan FA;
PI Thangavelu K, Thornton M, Lu DAM, Tribouley CM, Warren BA, Ison CH;
PI Das D, Raumann BE, Policky JL, Kearney L;

XX

DR WPI; 2002-463570/49.

DR N-PSDB; ABK83218.

XX

PT New transporters and ion channels (TRICH) polypeptides, useful for
PT diagnosing, preventing, and treating disorders associated with an
PT abnormal expression or activity of TRICH, e.g. immunological, muscular or
PT renal disorders.

XX

PS Claim 1; Page 143-144; 178pp; English.

XX

CC The invention relates to human transporters and ion channels (TRICH)
CC polypeptides, a naturally occurring amino acid sequence 90 % identical to
CC TRICH, a biologically active fragment of TRICH or an immunogenic fragment
CC of TRICH. Also included are an isolated polynucleotide encoding TRICH, a
CC recombinant polynucleotide comprising a promoter sequence operably linked
CC to the TRICH polynucleotide, a cell transformed with the recombinant
CC polynucleotide, a transgenic organism comprising the recombinant
CC polynucleotide, an isolated antibody that binds specifically to TRICH,
CC and screening for compounds which bind to TRICH, modulate TRICH, modulate
CC TRICH expression or are ant/agonists of TRICH. The polypeptides are
CC useful for diagnosing, treating, and preventing transport, neurological,
CC muscle, immunological disorders (e.g. scleroderma, systemic lupus
CC erythematosus, allergies), cell proliferative disorders such as cancers
CC (e.g. leukaemia, cervical or breast cancers), neurodegenerative disorders
CC (e.g. Parkinson's disease, Alzheimer's disease), muscular disorders (e.g.
CC myotonic dystrophy, catatonia), endocrine disorders (e.g. diabetes,
CC Grave's disease), gastrointestinal disorders (e.g. Crohn's disease),
CC renal disorders (e.g. Good pasture's syndrome), viral, bacterial, fungal,
CC parasitic, protozoal and helminthic infections, cardiovascular disorders
CC (e.g. atherosclerosis), or hepatic diseases (e.g. cirrhosis) and many
CC other diseases and disorders detailed in the specification. They can also
CC be used in assessing the effects of exogenous compounds on the expression
CC of nucleic acid and amino acid sequences of transporters and ion
CC channels. TRICH or its fragments may also be used in screening for
CC compounds that specifically bind to and modulate the activity of TRICH.
CC The polynucleotides can be used to create knock-in humanised animals or
CC transgenic animals to model human disease. The present sequence
CC represents a TRICH protein

XX

SQ Sequence 374 AA;

Query Match 55.9%; Score 1961; DB 5; Length 374;

Best Local Similarity 99.7%; Pred. No. 1.8e-192;

Matches 373; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 300 MVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDF 359
|| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

Db          1 MVHYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDF 60
Qy          360 LWKAETKDLDEDTCEVSSVTPLDTNCLPSP TKMPGAVQQFTTLIRRQISNDFRDLPTLLI 419
              |||
Db          61 LWKAETKDLDEDTCEVSSVTPLDTNCLPSP TKMPGAVQQFTTLIRRQISNDFRDLPTLLI 120
Qy          420 HGAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKYSERAMLY 479
              |||
Db          121 HGAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKYSERAMLY 180
Qy          480 YELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVWLVV 539
              |||
Db          181 YELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVWLVV 240
Qy          540 FCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWC 599
              |||
Db          241 FCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWC 300
Qy          600 FEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYY 659
              |||
Db          301 FEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYY 360
Qy          660 VSLRFIKQKPSQDW 673
              |||
Db          361 VSLRFIKQKPSQDW 374

```

RESULT 5

AAG18079

ID AAG18079 standard; protein; 632 AA.

XX

AC AAG18079;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 19344.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR	08-APR-1999;	99US-0128714P.
PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.
PR	23-APR-1999;	99US-0130510P.
PR	23-APR-1999;	99US-0130891P.
PR	28-APR-1999;	99US-0131449P.
PR	30-APR-1999;	99US-0132048P.
PR	30-APR-1999;	99US-0132407P.
PR	04-MAY-1999;	99US-0132484P.
PR	05-MAY-1999;	99US-0132485P.
PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	24-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136392P.
PR	28-MAY-1999;	99US-0136782P.
PR	01-JUN-1999;	99US-0137222P.
PR	03-JUN-1999;	99US-0137528P.
PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	16-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.

PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.

PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.

Query Match 20.8%; Score 730.5; DB 3; Length 632;
Best Local Similarity 30.7%; Pred. No. 2.1e-65;
Matches 211; Conservative 117; Mismatches 269; Indels 91; Gaps 19;

RESULT 6
AAG18078

ID AAG18078 standard; protein; 648 AA.
XX
AC AAG18078;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 19343.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.

PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	16-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.

PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.

PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161992P.
 PR 28-OCT-1999; 99US-0161993P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 20.8%; Score 730.5; DB 3; Length 648;
 Best Local Similarity 30.7%; Pred. No. 2.2e-65;
 Matches 211; Conservative 117; Mismatches 269; Indels 91; Gaps 19;

Qy 9 RGLPKGATPQDTSGLQDRLFSSS--DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFE 66
 :||| : | | | :| :| | || : :| : |
 Db 19 QGLPDMSDTQSKSVLAFPTITSQPGLQMSMY-----PITLKFEVVYKVKI-----E 65

Qy 67 QLAQFKMPWTSPSCQNSCELGIQNLSEFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGK 126
 | :| | | : : : | | :| :| | : :| : | |
 Db 66 QTSQCMGSWKSKE-----KTILNGITGMVCPGEFLAMLGPSGSGKTTLLSALGGR--LSK 118

Qy 127 IKSGQIWINQPPSSPQLVRKCAHVVRQHNLPLNTVRETLAFIAQMRLPRTFSQAQRDK 186
 ||:: |||| | : | : | | : | | :| | | | :|| : : : :
 Db 119 TFSGKVMYNGQPFSGCIKRR-TGFVAQDDVLYPHLTVWETLFFTALLRLPSSLTRDEKAE 177

Qy 187 RVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSE 246
 | : |||| | :| : : | | :||| :||| :| || :| :||| ||||
 Db 178 HVDRVIAELGLNRCTNSMIGGPLFRGISGGEKKRVSIGQEMLINPSLLLLDEPTSGLDST 237

Qy 247 TAHNLVKTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTA 306
 ||| :| | : ||| | | : :||| | | :|| :||| ||| | :|| : :
 Db 238 TAHRIVTTIKRLASGGRTVVTIHHQPSRIYHMFQKVVLLSEGSPIYYGAASSAVEYFSS 297

Qy 307 IGYPCPRYSNPADFYVDLTS-----IDRRSREQELATREKAQSLAALFLEKVRDLDDFLW 361
 :| : |||| :|| : : : ||| | : :|| : : : :
 Db 298 LGFSTSLTVNPADLLLDLANGIPPDTQKETSEQEQTQVK--ETLVSAIEKNI----- 347

Qy 362 KAETKDLDEDTCVES-----VTPLDTNCLPSPTKMPGAVQQFTTLIRRQI-SNDFRDLPT 416
 || | : | | | | | | | | : : | : |
 Db 348 --STK-LKAELCNAESHSEYTKAAAKNLKSEQWCTTWWYQFTVLLQRGVRRERREFESFNK 404
 Qy 417 LLIHGAEACLMSMTIGFLYFGHSGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSER 476
 | | : : | | : : | : : | | | | : : : | :
 Db 405 LRIF---QVISVAFLGGLLWWH-TPKSHIQDRTALLFFFSVFWGFYPLYNVETFFPQEK 460
 Qy 477 MLYYELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVW 536
 || | | : | | | : : | | | : | | : | : | : |
 Db 461 MLIKERSSGMYRLSSYFMARNVGDLPLELALPTAFVFIIYWMGGLKPDPTTFILSLIVVL 520
 Qy 537 LVVFCCRIMALAAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVP---AWISKV 593
 | : : | | | | : : | : | : : : | : :
 Db 521 YSVLVAQGLGLAFGALLMNIKQATTLASVTTLVFLIAGGYVQ-----QIPPFIVWLKYL 575
 Qy 594 SFLRWCFEGLMKIQFSRRTY-----KMPLGNLTIAVSGDKILSAMEL 635
 | : : | : | : | : | | | | | : : |
 Db 576 SYSYYCYKLLGLGIQYTDDDYYECSKGVWCRVGDFFPAIKSMGLNNLWI----DVFVMGVML 631
 Qy 636 DSYPLYAIYLIVIGLSGGFMVLYYVSLR 663
 | | | : | | | |
 Db 632 VGYRLMA-----YMALHRVKLR 648

RESULT 7

AAG18080

ID AAG18080 standard; protein; 625 AA.

XX

AC AAG18080;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 19345.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.
PR	23-APR-1999;	99US-0130510P.
PR	23-APR-1999;	99US-0130891P.
PR	28-APR-1999;	99US-0131449P.
PR	30-APR-1999;	99US-0132048P.
PR	30-APR-1999;	99US-0132407P.
PR	04-MAY-1999;	99US-0132484P.
PR	05-MAY-1999;	99US-0132485P.
PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	24-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136392P.
PR	28-MAY-1999;	99US-0136782P.
PR	01-JUN-1999;	99US-0137222P.
PR	03-JUN-1999;	99US-0137528P.
PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	16-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.

PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.

PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

Query Match 20.7%; Score 724; DB 3; Length 625;
Best Local Similarity 30.9%; Pred. No. 9.8e-65;
Matches 211; Conservative 113; Mismatches 258; Indels 100; Gaps 20;

```
Qy      15 ATPQDTS--GLQDRLFSSSEDNSLYFTYSQGPNTLEVRDLNYQVDLASQVPWFEEQLAQFK 72
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      11 AFPTITSQPGLQ-----MSMY-----PITLKFEENVYKVKI-----EQTSQCM 48

Qy      73 MPWTSPSCQNSCELGIQNLSEFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGGIKSGQI 132
      | | : : : : | | : | : | | | | : : | | : | | | | : :
Db      49 GSWKSKE-----KTILNGITGMVCPGEFLAMLGPSGSGKTLLSALGGR--LSKTFSGKV 101

Qy     133 WINGQPSSPQLVRKCAHVHRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVI 192
      | | | | | : | : | | : | | : | | | | | | | | : : : : : | : | |
Db     102 MYNGQPFSGCIKRR-TGFVAQDDVLYPHLTVWETLFFTALLRLPSSLTRDEKAEHVDRVI 160

Qy     193 AELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLV 252
      | | | | : | : : : | | : | | : | | : | | : | | : | | : | | : | |
Db     161 AELGLNRCNNSMIGGPLFRGISGGEKKRVSIGQEMLINPSLLLLDEPTSGLDSTTAHRIV 220

Qy     253 KTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCP 312
      | : | | | | | : : : | | : : | | : : : | : | | | | | | | : | : : : | :
Db     221 TTIKRLASGGRTVVTTIHQPSSRIYHMFQKVVLLSEGSPIYYGAASSAVEYFSSLSGFSTS 280

Qy     313 RYSNPADFYVDLTS-----IDRRSREQELATREKAQSLAALFLEKVRDLDDFLWKAETKD 367
      | | | | : | : : : : : : | | : : : | : : : : : | |
Db     281 LTVNPADLLLDLANGIPDPTQKETSEQEQKTVK--ETLVSAYEKNI-----STK- 327

Qy     368 LDEDTCEVSS----VTPLDTNCLPSPTKMPGAVQQFTTLIRRQI-SNDFRDLPTLLIHGA 422
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     328 LKAELCNAESHSEYETKAAAKNLKSEQWCTTWYQFTVLLQRGVRRERREFESFNKLRIF-- 385

Qy     423 EACLMSMTIGFLYFGHSGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYYEL 482
      : : : | : | : : : | | | | | | | : : : : | : | | |
Db     386 -QVISVAFLGGLLWWH-TPKSHIQDRTALLFFFSVFWGFYPLYNVAVFTFPQEKRMILIKER 443

Qy     483 EDGLYTTGPFYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVVLVVFCC 542
      | : | | | | : : : | | : : | | : | | : | | : | | |
Db     444 SSGMYRLSSYFMARNVGDLPLELALPTAFVFIIYWMGGLKPDPTTFILSLLVVLYSVLVA 503

Qy     543 RIMALAAAALLPTFHMAFFSNALYNSFYLAGGFMINLSSLTVP---AWISKVSFLRWC 599
      : : | | | | | : : : | : | | : : : : | : : | : : |
Db     504 QGLGLAFGALLMNIKQATTLASVTTLVFLIAGGYVQ-----QIPPFIVWLKYLSYSYYC 558

Qy     600 FEGLMKIQFSRRTY-----KMPLGNLTIAVSGDKILSAMELDSYPLY 641
      : : | : | : : | | | | | | | : : | | | |
Db     559 YKLLLGIIQYTDDDYYECSKGVWCRVGFPAIKSMGLNNLWI----DVFMGVMLVGYRLM 614

Qy     642 AIYLIVIGLSGGFMVLYYVSLR 663
      | : | : | | |
Db     615 A-----YMALHRVKLR 625
```

RESULT 8
AAU96986
ID AAU96986 standard; protein; 652 AA.

XX
 AC AAU96986;
 XX
 DT 07-AUG-2003 (revised)
 DT 30-JUL-2002 (first entry)
 XX
 DE Rat ABCG5 protein.
 XX
 KW Rat; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 XX
 OS Rattus sp.
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US029859.
 XX
 PR 25-SEP-2000; 2000US-0235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 DR N-PSDB; ABK51686.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Example 3; Page 45; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the rat ABCG5 protein of the invention. (Updated
 CC on 07-AUG-2003 to correct OS field.)

XX

SQ Sequence 652 AA;

Query Match 20.3%; Score 713; DB 5; Length 652;
Best Local Similarity 30.0%; Pred. No. 1.4e-63;
Matches 190; Conservative 115; Mismatches 232; Indels 96; Gaps 15;

```
Qy      12 PKGAT-PQDTSGLQDRLFSSSESDNSLYFTYSGQPNTLEVRDLNYQVDLASQV-PWFEQLA 69
      ||| | : | | || : | : | | : | | || : : : | ||
Db      9 PEGARGPHNNRGSQ-----SSLEEGSV--TGSEARHSLGV--LNVSFVSNSNRVGPW----- 55

Qy      70 QFKMPWTSPSCQNSCELGI-QNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIK 128
      | ||| : | : : | : ||| : | : ||| | : : ||| | : ||
Db     56 -----WNIKSCQQKWDKILKDVSLYIESGQTMCILGSSGSGKTLLDAISGRLRRTGTL 110

Qy     129 SGQIWINGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETIAFIAQMRLPRTFSQAQRDKRV 188
      | : : : || : || : : | : | : ||||| | : | : | | : | || : |
Db     111 EGEVFNVCCELRRDQFQDCVSYLLQSDVFLSSLTVRETILRYTAMLAL-RSSSADFYDKKV 169

Qy     189 EDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTFA 248
      | | : || | || : || | : | ||||| ||| : | : : ||||| : ||| ||
Db     170 EAVLTELSLSHVADQMIGNYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTA 229

Qy     249 HNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLMTSGTPPIYLGAQHMVQYFTAIG 308
      : : : | | || : || : : : ||||| : | || : : | | : | : | : | |
Db     230 NHIVLLLVELARRNRIVIVTIHQPRSELFHHFDKIAILTYGELVFCGTPEEMLGFFNNCG 289

Qy     309 YPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDL 356
      |||| : || ||| : |||| : | : ||| : | : : | | : | : | : | |
Db     290 YPCPEHSNPFDFYMDLTSVDTSREREIETYKRVQMLESAFRQSDICHKILENIERTRHL 349

Qy     357 DDFLWKAETKDLDEDTCESSVTPLDTNCLPSPTK-MPGAVQQFTTLIRRQISNDFRDLP 415
      | : : | || || : | : || | | :
Db     350 -----KTLP-----VPEKTKNPPGMFCKLGVLLRRVTRNLMRNKQ 385

Qy     416 TLLIHGAEACLSMTIGF--LYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCY 473
      : : : : | : : | | : : : | || : : | : : : :
Db     386 VVIMRLVQNLIMGLFLIFYLLRVQNMLKGAVQDRVGLLYQLVGATPYTGMLNAVNLFPM 445

Qy     474 ERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLLHFL 533
      || : | : ||| | : | || : : || | | : |
Db     446 LRAVSDQESQDGLYQKWQMLLAYVLHALPF SIVATVIFSSVCYWTGLGYPEVARF----- 500

Qy     534 LVWLVVFCRIMALAAAALLPTFHMASFFSNAL-----YNSFYLAGG 575
      : |||| : | : | : : |
Db     501 -----GYFSAALLAPHLIGEFLTLVLLGMVQNPNIIVNSIVALLSISGLLIGSG 548

Qy     576 FMINLSSLWTVPAWISKVSFLRWCFEGLMKIQF 608
      | : | : : : | : | | | : |
Db     549 FIRNIEEMPIPLKILGYFTFQKYCCEILVVNEF 581
```

RESULT 9

AAU96990

ID AAU96990 standard; protein; 651 AA.

XX

AC AAU96990;

XX
 DT 30-JUL-2002 (first entry)
 XX
 DE Human ABCG5 mutant R389H protein sequence.
 XX
 KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 389
 FT /note= "Wild-type Arg substituted by His"
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US029859.
 XX
 PR 25-SEP-2000; 2000US-0235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Claim 7; Page; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino

CC acid sequence represents the human ABCG5 mutant R389H protein of the
 CC invention. Note: This sequence is not shown in the specification but is
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages
 CC 35-36 of the specification
 XX
 SQ Sequence 651 AA;

Query Match 20.1%; Score 705; DB 5; Length 651;
 Best Local Similarity 29.0%; Pred. No. 9.5e-63;
 Matches 188; Conservative 124; Mismatches 240; Indels 96; Gaps 16;

Qy	16	TPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWF EQLAQFKMPW	75
		: : : : : : : : :	
Db	8	TPGGSMLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQQW	64
Qy	76	TSPSCQNSCELGIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI	134
		: : : : : : : : : : :	
Db	65	TRQI-----LKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTF-LGEVYV	115
Qy	135	NGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAE	194
		: : : : : : : : : : : :	
Db	116	NGRALRREQFQDCFSYVLQSDTLLSSLTVRET LHYTALLAI-RRGNPGSFQKKVEAVMAE	174
Qy	195	LRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPTSGLDSFTAHNLVKT	254
		: : : : : : : : : :	
Db	175	LSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVM LFDEPTTGLDCMTANQIVVL	234
Qy	255	LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTP IYLGAAQHMVQYFTAIGYPCPRY	314
		: : : : : : : : : : :	
Db	235	LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEH	294
Qy	315	SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK	362
		: : : : : : : : : :	
Db	295	SNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS AICHKTLKNIERMKHL-----	348
Qy	363	AETKDLDEDTCESSVTPLDTNCLPSPTK-MPGAVQQFT TLRQISNDFRDLPTLLIHG	421
		: : : : : : : : : : : :	
Db	349	-----KTLPM----VPFKTKDSPGVFSKLGVL LRRVTRNLVRNKLAVITHL	390
Qy	422	AEACILMSMTIGFLYFG-----HGSIQLSFMDTAALL FMIGALIPFNVILDVISKYSER	475
		: : : : : : : : : : : :	
Db	391	LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DRVG LLYQFVGATPYTGMLNAVNLPVLR	446
Qy	476	AMLYYELEDGLYTTGPHYFFAKILGELPEHCAYII IYGMPTYWLANLRPGLQPFLHFLLV	535
		: : : : : : : : :	
Db	447	AVSDQESQDGLYQKWQMLLAYALHVL PFSV VATMIFSSVCYWTGLHPEVARF-----	499
Qy	536	WLVVFCCRIMALAAAALLPTFHMAFFS-----NALY NSFYLAG-----GFM	577
		: : : : : : : : : :	
Db	500	-----GYFSAALLAPHLIGEFLT LVLLGIVQNPNI VNSVALLSIAGVLVSGSFL	549
Qy	578	INLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTY KMPLGNLTI AVS	625
		: : : : : : : : : : :	
Db	550	RNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFT CGSSNVSVTTN	597

AAU96993

ID AAU96993 standard; protein; 651 AA.

XX

AC AAU96993;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R419P protein sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW mutant; mutein.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 419

FT /note= "Wild-type Arg substituted by Pro"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Claim 10; Page; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.

RESULT 11

AAU96984

ID AAU96984 standard; protein; 651 AA.

XX

AC AAU96984;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 protein.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW chromosome 2p21.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Misc-difference 2. .15

FT /note= "Encoded by GGTCTC"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR N-PSDB; ABK51681.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic

PT acid encoding the polypeptide, useful for treating sitosterolemia,

PT arteriosclerosis and heart diseases.

XX

PS Claim 52; Page 35-36; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying a
CC compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal

CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 protein of the invention. This
 CC sequence is encoded by the human ABCG5 gene located on chromosome 2p21
 XX
 SQ Sequence 651 AA;

Query Match 19.9%; Score 697; DB 5; Length 651;
 Best Local Similarity 28.9%; Pred. No. 6.3e-62;
 Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

Qy	16	TPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWF EQLAQFKMPW	75
		: : : : : : : : :	
Db	8	TPGGSMGLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQQW	64
Qy	76	TSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI	134
		: : : : : : : : : : : : :	
Db	65	TRQI-----LKDVSLYVESGQIMCILGSSGSGKT'TLLDAMSGRLGRAGTF-LGEVYV	115
Qy	135	NGQPSSPQLVRKCVAHVRQHNLPLNTVRET LAFIAQMRLPRTFSQAQRDKRVEDVIAE	194
		: : : : : : : : : : : : :	
Db	116	NGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAI-RRGNPGSFQKKVEAVMAE	174
Qy	195	LRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPTSGLDSFTAHLNVKT	254
		: : : : : : : : : : :	
Db	175	LSLSHVADRLIGNYSLGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVL	234
Qy	255	LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY	314
		: : : : : : : : : : : :	
Db	235	LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFNDCGYPCPEH	294
Qy	315	SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK	362
		: : : : : : : : : : :	
Db	295	SNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS AICHKTLKNIERMKHL-----	348
Qy	363	AETKDLDEDTCVESSVTPLDTNCLPSP TK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHG	421
		: : : : : : : : : : :	
Db	349	-----KTLPM----VPFKTKDSPGVFSKLGVL LRRVTRNLVRNKLAVITRL	390
Qy	422	AEACLMSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSER	475
		: : : : : : : : : : : : :	
Db	391	LQNLIMGLFLLFFVLRVRSNVLKGAIQ---DRVGLLYQFVGATPYTGMLNAVNLFPPVLR	446
Qy	476	AMLYYELEDGLYTTGPYFFAKILGELPEHCAYII IYGMPTYWLANLRPGLQPFL LHFLLV	535
		: : : : : : : :	
Db	447	AVSDQESQDGLYQKWQMMLAYALHVL PFSVVATMIFSSVCYWTGLGHPEVARF-----	499
Qy	536	WLVVFCCRIMALAAAAALLPTFHMA SFFS-----NALYNSFYLAG-----GFM	577
		: : : : : : : : : :	
Db	500	-----GYFSAALLAPHLIGEFLTLVLLGIVQNPNI VNSVALLSIAGVLVGSGFL	549

Qy 578 INLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVS 625
 |: : || :| :| | |: :| : |:: :
 Db 550 RNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTN 597

RESULT 12

AAE13290

ID AAE13290 standard; protein; 651 AA.

XX

AC AAE13290;

XX

DT 12-FEB-2002 (first entry)

XX

DE Human sitosterolaemia susceptibility gene (SSG) protein.

XX

KW Human; sitosterolaemia susceptibility gene; SSG; atherosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21.

XX

OS Homo sapiens.

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.

XX

PI Tian H, Schultz J, Shan B;

XX

DR WPI; 2002-017598/02.

DR N-PSDB; AAD22009.

XX

PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,
 PT useful for screening a compound that increases the level of expression or
 PT activity of SSG polypeptide for treating sterol-related disorder.

XX

PS Claim 19; Fig 8; 105pp; English.

XX

CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful for
 CC identifying a compound useful in the treatment or prevention of a sterol-
 CC related disorder, including sitosterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolaemia and other

AAE31704

ID AAE31704 standard; protein; 651 AA.

XX

AC AAE31704;

XX

DT 24-MAR-2003 (first entry)

XX

DE Human ABCG5 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;

KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;

KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;

KW ABCG5.

XX

OS Homo sapiens.

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.

XX

PF 20-NOV-2001; 2001WO-US043823.

XX

PR 20-NOV-2000; 2000US-0252235P.

PR 28-NOV-2000; 2000US-0253645P.

XX

PA (TULA-) TULARIK INC.

PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Hobbs HH, Shan B, Barnes R, Tian H;

XX

DR WPI; 2003-058548/05.

DR N-PSDB; AAD48882.

XX

PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-

PT related disorders e.g. sitosterolemia, hypercholesterolemia,

PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or

PT nutritional deficiencies.

XX

PS Claim 28; Page 78-79; 94pp; English.

XX

CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known
CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention
CC are useful for treating or preventing sterol-related disorders such as
CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL
CC deficiency, atherosclerosis and nutritional deficiencies. They are also
CC useful in gene therapy. The present sequence is human ABCG5 protein

XX

SQ Sequence 651 AA;

Query Match 19.9%; Score 697; DB 6; Length 651;

Best Local Similarity 28.9%; Pred. No. 6.3e-62;

Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

Qy 16 TPQDTSGLQDRLESSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWF EQLAQFKMPW 75

```

      || : ||| | | | : : || : : | : || : || : : |
Db      8 TPGGSMGLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQOW 64

Qy      76 TSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI 134
      | : : : | | ||| : | ||| | : ||| : || | | : : :
Db      65 TRQI-----LKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTF-LGEVYV 115

Qy      135 NGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAE 194
      || : : : | : | | : ||||| : | : : | : | : || : ||
Db      116 NGRALRREQFDQCFSYVLQSDTLLSSLTVRETLHYTALLAI-RRGNPGSFQKKVEAVMAE 174

Qy      195 LRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKT 254
      | | || : || : | : | ||||| ||| : | : : |||| : || | : : |
Db      175 LSLSHVADRLIGNYSLGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVL 234

Qy      255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314
      | || : || : : : |||| : : || : : : | | : | | : : | |||| :
Db      235 LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEH 294

Qy      315 SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDDFLWK 362
      ||| ||| : ||| : | : | : | : : | : : : | : : : |
Db      295 SNPDFFYMDLTSVDTSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHL----- 348

Qy      363 AETKDLDEDTCESSVTPLDTNCLPSPTK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHG 421
      | : : : | || || : | : || | | : :
Db      349 -----KTLPM----VPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRL 390

Qy      422 AEACLSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSER 475
      : : | : : | | : || | || : | : : : : |
Db      391 LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DRVGLLYQFVGATPYTGMLNAVNLFVPLR 446

Qy      476 AMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLV 535
      | : | : ||| | | || : : || | | : |
Db      447 AVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLGLHPEVARE----- 499

Qy      536 WLVVFFCCRIMALAAAALLPTFHMAFFS-----NALYNSFYLAG-----GFM 577
      : ||| : | : | : : || : || :
Db      500 -----GYFSAALLAPHLIGEFLTLVLLGIVQNPNIVNSVVALLSIAGVLVSGGFL 549

Qy      578 INLSSLWTPPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVS 625
      | : : || : | : | | : : | : : : :
Db      550 RNIQEMPIPFKIISYFTFQKYCSEILVNNEFYGLNFTCGSSNVSVTTN 597

```

RESULT 14

AAU96989

ID AAU96989 standard; protein; 651 AA.

XX

AC AAU96989;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R419H protein sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW mutant; mutein.

XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 419
 FT /note= "Wild-type Arg substituted by His"
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US029859.
 XX
 PR 25-SEP-2000; 2000US-0235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Claim 9; Page; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R419H protein of the
 CC invention. Note: This sequence is not shown in the specification but is
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages
 CC 35-36 of the specification
 XX
 SQ Sequence 651 AA;

Query Match 19.9%; Score 696; DB 5; Length 651;

Best Local Similarity 28.9%; Pred. No. 8e-62;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

```

Qy      16 TPQDTSGLDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFEOQLAQFKMPW 75
      || : ||| | | : : ||: | : | : ||: : : |
Db      8 TPGGSMGLQVNRGSQS SLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQOW 64

Qy      76 TSPSCQNSCELGIQNLSFKVRSQM LAIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI 134
      | : : : | | |||: | : ||| | : || | | : : :
Db      65 TRQI-----LKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTF-LGEVYV 115

Qy     135 NGQPSSPQLVRKCVAHVRQHNLPLNTVRET LAFIAQMRLPRTFSQAQRDKRVEDVIAE 194
      ||: : : | : | | : ||||| | : | : | : | : || | : ||
Db     116 NGRALRREQFQDCFSYVLQSDTLLSSLTVRET LHYTALLAI-RRGNPGSFQKKVEAVMAE 174

Qy     195 LRLRQCADTRVGNMYVRGLSGGERRRV SIGVQLLWNPGLIILDEPTSGLD SFTAHLVKT 254
      | | || : || : | : | ||||| ||| : | : : |||: ||| ||: |
Db     175 LSLSHVADRLIGNYSLGGISTGERRRV SIAAQLLQDPKVM LFDEPTTGLDCMTANQIVVL 234

Qy     255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314
      | ||: ||: ||: : ||||: ||: ||| : : : | | : | | : | |||| :
Db     235 LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFND CGYPCPEH 294

Qy     315 SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK 362
      ||| |||: |||: | : | : | : : | : : : : | : : : |
Db     295 SNPFD FYMDLTSVDTQSKEREIETSKRVQMIESAYKKS AICHKTLKNIERMKHL----- 348

Qy     363 AETKDLDEDTCESSVTPLDTNCLPSP TK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHG 421
      | : : | || || : | : || | | : :
Db     349 -----KTLPM----VPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRL 390

Qy     422 AEACLMSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSER 475
      : : | : : | | : || | ||: | : : | : : : |
Db     391 LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DHVG LLYQFVGATPYTGMLNAVNLFVPLR 446

Qy     476 AMLYYELEDGLYTTGPFYFFAKILGELPEHCAYII IYGMPTYWLANLRPGLQPFL LHFLLV 535
      | : | : |||| | | || : : || | | : |
Db     447 AVSDQESQDGLYQKWQMLLAYALHVL PFSVVATMIFSSVCYWT LGLHPEVARF----- 499

Qy     536 WLVVFFCCRIMALAAAALLPTFHMASFFS-----NALYNSFYLAG-----GFM 577
      : |||| : | : | : : : || ||:
Db     500 -----GYFSAALLAPHLIGEFLTLVLLGIVQNP NIVNSVVALLSIAGVLVGS GFL 549

Qy     578 INLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVS 625
      | : : || : | : || | : | : : | : : :
Db     550 RNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTN 597

```

RESULT 15

AAU96992

ID AAU96992 standard; protein; 651 AA.

XX

AC AAU96992;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant E146Q protein sequence.

XX
 KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 146
 FT /note= "Wild-type Glu substituted by Gln"
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US029859.
 XX
 PR 25-SEP-2000; 2000US-0235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR WPI; 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic
 PT acid encoding the polypeptide, useful for treating sitosterolemia,
 PT arteriosclerosis and heart diseases.
 XX
 PS Claim 12; Page; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying a
 CC compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant E146Q protein of the
 CC invention. Note: This sequence is not shown in the specification but is
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages
 CC 35-36 of the specification

XX

SQ Sequence 651 AA;

Query Match 19.8%; Score 694; DB 5; Length 651;
Best Local Similarity 28.7%; Pred. No. 1.3e-61;
Matches 186; Conservative 125; Mismatches 241; Indels 96; Gaps 16;

```
Qy      16 TPQDTSGLDRLFSSESNDLSLYFTYSGQPNLTLEVRDLNYQVDLASQVPWFQQLAQFKMPW 75
      || : ||| | | | : : || : : || : || : : |
Db      8 TPGGSMGLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQOW 64

Qy      76 TSPSCQNSCELGIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI 134
      | : : : | | || : | : || | : : || | | | : : :
Db      65 TRQI-----LKDVSLYVESGQIMCILGSSSGKTTLLDAMSGRLLGRAGTF-LGEVYV 115

Qy     135 NGQPSSPQLVRKCVAVHRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAE 194
      || : : : | : | | : || : || : || : | : : | : | : || | : ||
Db     116 NGRALRREQFQDCFSYVLQSDTLLSSLTVRQTLHYTALLAI-RRGNPGSFQKKVEAVMAE 174

Qy     195 LRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLNVKT 254
      | | || : || : | : | || || || || || : || : || : || : || : |
Db     175 LSLSHVADRLIGNYSLGISTGERRRVSIAAQLLQDPKVMLEFDEPTTGLDCMTANQIVVL 234

Qy     255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314
      | || : || : : : || || : : || : : : | : | : : | || || :
Db     235 LVELARRNRIVVLTIHQPRSELFQLFDKIALLSFGELIFCGTPAEMLDFFNDGYPCEH 294

Qy     315 SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK 362
      ||| ||| : ||| : | : | : | : : | : : : : : : : : : | : : : |
Db     295 SNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHL----- 348

Qy     363 AETKDLDEDTCESSVTPLDTNCLPSPTK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHG 421
      | : : : | | | | : | : || | | : :
Db     349 -----KTLPM---VPFKTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRL 390

Qy     422 AEACLSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVILDVISKYCYSER 475
      : : | : : | : || | | : || : | : : : : |
Db     391 LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DRVGLLYQFVGATPYTGMLNAVNLFPVLR 446

Qy     476 AMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLV 535
      | : | : ||| | | || : : || | | : |
Db     447 AVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLGLHPEVARF----- 499

Qy     536 WLTVFCCRIMALAAAALLPTFHMASFFS-----NALYNSFYLAG-----GFM 577
      : ||| : | : : : : | : : : || || :
Db     500 -----GYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSVALLSIAGVLVSGSGL 549

Qy     578 INLSSLWTPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAS 625
      | : : || : | : | | : | : : : :
Db     550 RNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTN 597
```

Search completed: February 27, 2004, 06:44:22
Job time : 49.0351 secs

OM protein - protein search, using sw model

Run on: February 27, 2004, 07:11:48 ; Search time 15.2492 Seconds
(without alignments)
2278.426 Million cell updates/sec

Title: US-09-989-981A-8
Perfect score: 3506
Sequence: 1 MAGKAAEERGLPKGATPQDT.....FMVLYYVSLRFIKQKPSQDW 673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	640.5	18.3	655	4	US-09-245-808-1	Sequence 1, Appli
2	638.5	18.2	655	4	US-09-767-594-1	Sequence 1, Appli
3	497.5	14.2	617	4	US-09-614-912-138	Sequence 138, App
4	481.5	13.7	1296	4	US-09-614-912-140	Sequence 140, App
5	440.5	12.6	539	4	US-09-614-912-144	Sequence 144, App
6	250	7.1	1684	3	US-08-665-259-25	Sequence 25, Appl
7	250	7.1	1684	3	US-08-762-500-25	Sequence 25, Appl
8	250	7.1	1704	3	US-08-762-500-75	Sequence 75, Appl
9	243.5	6.9	265	4	US-09-489-039A-10393	Sequence 10393, A
10	243	6.9	593	4	US-09-252-991A-21665	Sequence 21665, A
11	242	6.9	330	4	US-09-252-991A-27569	Sequence 27569, A

12	238	6.8	379	4	US-09-489-039A-11991	Sequence 11991, A
13	231	6.6	388	4	US-09-489-039A-8815	Sequence 8815, Ap
14	230	6.6	370	4	US-09-252-991A-20719	Sequence 20719, A
15	229	6.5	229	4	US-09-134-000C-3584	Sequence 3584, Ap
16	229	6.5	362	4	US-09-489-039A-12150	Sequence 12150, A
17	228.5	6.5	359	4	US-09-328-352-6329	Sequence 6329, Ap
18	227.5	6.5	347	4	US-09-543-681A-5961	Sequence 5961, Ap
19	227	6.5	432	4	US-09-328-352-6798	Sequence 6798, Ap
20	226.5	6.5	360	4	US-09-540-236-2029	Sequence 2029, Ap
21	225	6.4	345	4	US-09-252-991A-18872	Sequence 18872, A
22	223	6.4	344	4	US-09-540-236-2702	Sequence 2702, Ap
23	222.5	6.3	1457	3	US-08-665-259-27	Sequence 27, Appl
24	222.5	6.3	1457	3	US-08-762-500-27	Sequence 27, Appl
25	220.5	6.3	374	4	US-09-489-039A-10660	Sequence 10660, A
26	219.5	6.3	396	4	US-08-311-731A-209	Sequence 209, App
27	219.5	6.3	676	4	US-09-252-991A-21751	Sequence 21751, A
28	219	6.2	532	4	US-09-543-681A-4646	Sequence 4646, Ap
29	217.5	6.2	235	4	US-09-198-452A-242	Sequence 242, App
30	217.5	6.2	384	4	US-09-489-039A-9127	Sequence 9127, Ap
31	217	6.2	683	4	US-09-252-991A-26621	Sequence 26621, A
32	216.5	6.2	345	4	US-09-107-532A-3849	Sequence 3849, Ap
33	215	6.1	250	4	US-09-328-352-7153	Sequence 7153, Ap
34	215	6.1	256	4	US-09-134-001C-4600	Sequence 4600, Ap
35	215	6.1	372	4	US-09-489-039A-10875	Sequence 10875, A
36	214.5	6.1	376	2	US-08-997-080-89	Sequence 89, Appl
37	214.5	6.1	376	2	US-08-997-362-89	Sequence 89, Appl
38	214.5	6.1	376	3	US-08-873-970-89	Sequence 89, Appl
39	214.5	6.1	376	3	US-09-095-855-89	Sequence 89, Appl
40	214.5	6.1	376	4	US-09-324-542-89	Sequence 89, Appl
41	214.5	6.1	376	4	US-09-205-426-89	Sequence 89, Appl
42	213.5	6.1	373	4	US-09-543-681A-7638	Sequence 7638, Ap
43	213.5	6.1	383	4	US-09-252-991A-25307	Sequence 25307, A
44	213	6.1	300	4	US-09-489-039A-9673	Sequence 9673, Ap
45	213	6.1	360	4	US-09-543-681A-6466	Sequence 6466, Ap

ALIGNMENTS

RESULT 1

US-09-245-808-1

; Sequence 1, Application US/09245808

; Patent No. 6313277

; GENERAL INFORMATION:

; APPLICANT: Doyle, L. Austin

; APPLICANT: Abruzzo, Lynne V.

; APPLICANT: Ross, Douglas D.

; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

; TITLE OF INVENTION: encodes it

; FILE REFERENCE: Ross UMb conversion

; CURRENT APPLICATION NUMBER: US/09/245,808

; CURRENT FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/073763

; EARLIER FILING DATE: 1998-02-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 655
; TYPE: PRT
; ORGANISM: Human MCF-7/AdrVp cells
US-09-245-808-1

Query Match 18.3%; Score 640.5; DB 4; Length 655;
Best Local Similarity 27.2%; Pred. No. 5.1e-62;
Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

```
Qy      19 DTSG LQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWF EQLAQFKMPWTSP 78
          :|:|      | : : | |      |      :: :| | |      :|
Db      16 NTNG-----FPATASNDLKAFTEGA--VLSFHNICYRVKLKSGF-----LP----- 54

Qy      79 SCQNSCELGI-QNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQI WINGQ 137
          |:      | | |:: :| : ||:| :| |::||| |:      | : || : |||
Db      55 -CRKPVEKEILS NINGIMKPG-LNAILGPTGGGKSSLLDVL AARKDPSGL-SGDVLINGA 111

Qy     138 PSSPQLVRKC-VAHV RQHNQLLPNLT VRET LAFIAQMRLPRTFSQAQRDKRVEDVIAELR 196
          |      || :| | : :: |||| | | | :|| | : :::|: | |||
Db     112 PRPANF--KCN SGYVVQDDVVMGTLTVREN LQFS AALRLATMTNHEKNERINRVIQELG 169

Qy     197 LRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGILILDEPTSGLDSFTAHNLVKTLS 256
          | : ||::|| :||:||||:| ||:::| :| || ||||:|||| ||: : : |
Db     170 LDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSS TANA VLLLLK 229

Qy     257 RLA KGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSN 316
          |:| | : : |:|||| ||:|||| : | : || :| | : || : || | |:|
Db     230 RMSKQGR TIIFS I HQPRYSIFKLFD SLTLLASGRLMFHGPAQEALGYFESAGYHCEAYNN 289

Qy     317 PADFYVDLTSIDRR----SREQELATRE--KAQSLAALFLEKVRDL--DDFLWKAETK-- 366
          |||::|: : |      :||:| | :      :||: : : : :| |||
Db     290 PADFFLDIINGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLA EIYVNSSFYK-ETKAE 348

Qy     367 -----DLDEDTCESSVTPLDTNCLPSP TKMPGAVQQFTTLIR RQISNDFRDLPTLLI 419
          : : | : : : |      | : :| | : :
Db     349 LHQLSGGEKKKKITVFKEISYTT SFC-----HQLRWVSKRSFKNLLGNPQASIA 397

Qy     420 HGAEACLMSMTIGFLYFGHGS IQLSFMDTAALLFMIGALIPFNVILDVISKCY S----- 473
          :: : || :||| :      : | :|| :      :::|
Db     398 QIIVTVVLGLVIGAIYFGLKNDSTGIQNRAGVLF FL-----TTNQCFSSVSAVE 446

Qy     474 ----ERAMLYYELEDGLYTTGPYFFAKILGE-LPEHCAYIIIIYGMPTYWLANLRPGLQPF 528
          |: : :| | | || |:| : || ||: ||: |:| |
Db     447 LFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLPMTMLPSIIFT CIVYFMLGLKPKADAF 506

Qy     529 LLHFLLVWL VVFCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSS LWT VPA 588
          : : :| : ||| || :|: : : : | :||::: : :
Db     507 FVMFTLMMVAYSASSMALAIAAGQSVVSVATLLMTICFVFM MIFSGLLVNLTTIASWLS 566

Qy     589 WISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT-----IAVSGDKIL--SAMELDSYP 639
          |: | |:| | :| : : | | | :|| | :| :
Db     567 WLQYFSIPRYGFTALQHNEFLGQNF-CPGLNATGNNPCNYATCTGEEYLVKQGIDLSPWG 625

Qy     640 LYAIYLIVIGLSGGFMVLYYVSLRFIKQ 667
          |: : : : | : : |: | |:|
Db     626 LWKNHVALACMIVIFLTIA YLKLFLKK 653
```

RESULT 2
 US-09-767-594-1
 ; Sequence 1, Application US/09767594
 ; Patent No. 6521635
 ; GENERAL INFORMATION:
 ; APPLICANT: Bates, Susan
 ; APPLICANT: Robey, Robert
 ; APPLICANT: The Government of the United States of America
 ; APPLICANT: as represented by the Secretary of the
 ; APPLICANT: Department of Health and Human Services
 ; TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives
 ; FILE REFERENCE: 015280-402100US
 ; CURRENT APPLICATION NUMBER: US/09/767,594
 ; CURRENT FILING DATE: 2001-01-22
 ; PRIOR APPLICATION NUMBER: US 60/177,410
 ; PRIOR FILING DATE: 2000-01-20
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 655
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: human mitoxanthrone resistance (MXR)/BRCP/ABCP
 ; OTHER INFORMATION: protein
 US-09-767-594-1

Query Match 18.2%; Score 638.5; DB 4; Length 655;
 Best Local Similarity 27.9%; Pred. No. 8.5e-62;
 Matches 175; Conservative 131; Mismatches 254; Indels 67; Gaps 17;

Qy	80	CQNSCELGI-QNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQGP	138
		: :: :: : : : :: : :	
Db	55	CRKPVEKEILSNINGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPSGL-SGDVLINGAP	112
Qy	139	SSPQLVRKC-VAHVRQHNLQLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRL	197
		: : :: : : ::: :	
Db	113	RPANF--KNSGYVVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIEELGL	170
Qy	198	RQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLRS	257
		: :: : : :	
Db	171	DKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILSLDEPTTGLDSSTANAVLLLLKR	230
Qy	258	LAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNP	317
		:: : : : : : : : :	
Db	231	MSKQGRTIIFSIIHQPRYSIFKLFDSLTLASGRLMFHGPAQEALGYFESAGYHCEAYNNP	290
Qy	318	ADFYVDLTSIDRR----SREQELATRE--KAQSLAALFLEKVRDL--DDFLWKAETK---	366
		:: : : : : : : : : : : : :	
Db	291	ADFFLDIINGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYK-ETKAEL	349
Qy	367	-----DLDEDTCVESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH	420
		: : : : : : : : : : : : : :	
Db	350	HQLSGGEKKKKITVFKEISYTTSF-----HQLRWVSKRSFKNLLGNPQASIAQ	398
Qy	421	GAEACLMSMTIGFLYFGHSGIQLSFMDTAALLFMIGALIPFNVILDVISKYS-----	473

```

      : : | | : | | :      : | : | | :      : : | |
Db      399 IIVTVVLGLVIGAIYFGLKNDSTGIQNRAGVLFLL-----TTNQCFSSVSAVEL 447

Qy      474 ---ERAMLYYELEDGLYTTGPYFFAKILGE-LPEHCAYIIIIYGMPTYWLANLRPGLQPFL 529
      | : : : | | | | | | | : | : | | | | : | : | |
Db      448 FVVEKKLFIHEYISGYRVSSYFLGKLLSDDLPMRMLPSIIFTCIVYFMLGLKPKADAFF 507

Qy      530 LHFLLVVLVVFCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAW 589
      : : : | : | | | | | | : | : : : | : | : | : : |
Db      508 VMMFTLMMVAYSASSMALAIAAGQSVSVATLLMTICFVMMIFSGLLVNLTTIASWLSW 567

Qy      590 ISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT-----IAVSGDKIL--SAMELDSYPL 640
      : | | : | | : | : : | | | : | : | : | : | : |
Db      568 LQYFSIPRYGFTALQHNEFLGQNF-CPGLNATGNNPCNYATCTGEEYLVKQGIDLSPWGL 626

Qy      641 YAIYLIVIGLSGGFMVLYYVSLRFIKQ 667
      : : : : | : : | : | | : | :
Db      627 WKNHVALACMIVIFLTIAYLKLLFLKK 653

```

RESULT 3

```

US-09-614-912-138
; Sequence 138, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 138
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Zea mays

```

US-09-614-912-138

Query Match 14.2%; Score 497.5; DB 4; Length 617;
Best Local Similarity 26.6%; Pred. No. 4.5e-46;
Matches 179; Conservative 125; Mismatches 246; Indels 123; Gaps 28;

```
Qy      44 PNTLEVRDLNYQVDLASQVPWFELAQFKMPWTSPSCQNSCELGIQN-----LSFKV 95
      | :  :|| ||: ::  |::  ::
Db      8 PLAMSFDNVNYVDMPEAMK-----HQGVQDDRLQLLREVTGSF 46

Qy     96 RSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSSPQLVRKCVAHVRQHN 155
      | | : |::| || | : |::| | | | | :  :  : |::
Db     47 RPGVLTALMGVSGAGKTTLMVDLAGRKTTGGYIE-GDIRIAGYPKNQATFARISGYCEQND 105

Qy    156 QLLPNLTVRETAFIAQMRLPRTFSQAQ-----RDKRVEDVIAELRLRQCADTRVGNMYV 210
      | :|||:| : | :||  :  : : |::| : | | :
Db    106 IHSPQVTVRESLIYSAFLRLPGKIGDQEITDDIKMQFVDEVMELVELDNLRDALVGLPGI 165

Qy    211 RGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTL SRLAKGNRLVLISLH 270
      ||| :|::| | :| : || | : |::| | | : | : : | | : : |
Db    166 TGLSTEQRKRLTIAVELVANPSIIFMDEPTSGLDARAAAIMRTVRNTVDTGRTVVCTIH 225

Qy    271 QPRSDIFRLFD-LVLLMTSGTPIYLGA----AQHMQYFTAI-GYP--CPRYSNPADFYV 322
      || ||| || |::| | || | : |::| | | | : | ||| : :
Db    226 QPSIDIFESFDELLLLKRGQVIYSGKLGRNSQKMVEYFEAIPGVPKIKDKY-NPATWML 284

Qy    323 DLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFLWKAETKDLDEDTCV---ESSVT 379
      ::|:  ||  :||  ||  || || :  | : |
Db    285 EVSSV-----AT-----EVRLKMDFAKYYETSDLYKQNKVLVNQLSQP 322

Qy    380 PLDTNCLPSPK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHGAELMSMTIGFLYFGH 438
      | : | ||:  : || : :| :|  | : : |:: : | ::
Db    323 EPGTSDLYFPTEYSQSTIGQFKACLWKQWLTYWRSPDYNLVRYSFLLVALLLGSIFWRI 382

Qy    439 GSIQLSFMDTAALLEFMIGALIPFNVILDV-ISKCY-----ERAMLYYELEDGLYTTG 490
      | :  : | | :||| : : : :| : | : || : | | | :| :
Db    383 GT---NMEDATTLGMVIGAM--YTAVMFIGINNCSTVQPVVSIERTVFYRERAAGMYSAM 437

Qy    491 PYFFAKILGELP----EHCAI-IIIYGMPTY-WLANLRPGLQPFLHFLLVWLVVFCRI 544
      || |::: |::  :  | :|:| | : : | |  | | : :
Db    438 PYAIAQVVIEIPYVVFQTTYTLIVYAMMSFQWTA-----VKFFWFFFISYFSFLYFTY 491

Qy    545 MALAAAALLPTFHMAFFSNALYNSFYLAGGFMI---NLSSLWTPPAWISKVSFLRWCFE 601
      : | :: | :|| | : : | | || | :  | ||  | |
Db    492 YGMAVSISPNEVASIFAAFFSLFNLFSGFFIPRPRIPGWWIYWWICP---LAWTVY 548

Qy    602 GLMKIQFSRRTYKMPLGNL--TIAVSG--DKILS-----AMELDSYPLYAIYLIVIGL 650
      ||: | :  | :| | :| | : : |  | | : | | : :
Db    549 GLIVTQY-----GDLEDLISVPGESEQTISYYVTHHFGYHRDFLPVIAPVLVLFVAV 599

Qy    651 SGGFMVLYYVSLR 663
      | || | : :
Db    600 F--FAFLYAVCIK 610
```

RESULT 4

US-09-614-912-140

```

; Sequence 140, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 140
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-614-912-140

```

```

Query Match          13.7%; Score 481.5; DB 4; Length 1296;
Best Local Similarity 25.5%; Pred. No. 1e-43;
Matches 180; Conservative 133; Mismatches 271; Indels 123; Gaps 29;

```

```

Qy      4 KAAEERGLPKGATPQDTSGLDRLFSSES DNS-----LYFTYSGQPNTLEVRDLNYQV 56
      | | | : :|: || | |: : | | | :| :| | |
Db      645 KEMREMRLSARLSNSSSNGV-SRLMSIGSNEAGPRRGMVLPFT----PLSMSFDDVNYV 699

Qy      57 DLASQVPWFELAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLD 116
      |: ::: : :| :::: | :|::| || |: :|:|
Db      700 DMPAEMK-----QQGVVDDRLLQL-LRDVTGSFRPAVL TALMGVSGAGKTTLMD 746

Qy      117 VITGRGHGGKIKSGQIWINQPPSSQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLP 176
      |: || || |: | :|:| | : : : :|: | :|::| | :|:| |
Db      747 VLAGRKTTGGYIE-GDMRISGYPKNQETFARISGYCEQNDIHS PQVTVRESLIYSAFLRLP 805

Qy      177 RTFSQAQ-----RDKRVEDVIAELRLRQCADTRVG NMYVRGLSGGERRRVSIGVQLLWNP 231
      : : :|::|: : | | | : || | :|:|:| |:|: | |
Db      806 EKIGDQEITDDIKIQFVDEVMELVELDNLKDALVGLPGITGLSTEQRKRLTI AVELVANP 865

```

Qy 232 GILILDEPTSGLDSTAHNLVKTLSRLAKGNRLVLISLHQPRSDIFRLFD-LVLLMTSGT 290
 | : : | | | | | : | : : : | : : : | | | | : | | |
 Db 866 SIIFMDEPTSGLDARAAIWMRTVRNTVDTGRTVVCTIHQPSIDIFEAFDELLLLKRGQ 925
 Qy 291 PIYLGA----AQHMQYFTAI-GYP--CPRYSNPADFYVDLTSIDRRSREQELATREKAQ 343
 | | | : | | : | | | | : | | | : : : : | :
 Db 926 VIYSGQLGRNSQKMIYFEAIPGVPKIKDKY-NPATWMLEVSSV----- 968
 Qy 344 SLAALFLEKVRDLDDFLWKAETKDLEDCTCV---ESSVTPLDTNCLPSPTK-MPGAVQQF 399
 : | | | : | | : | : | : | | | : | |
 Db 969 -----AAEVRLNMDFAEYYKTS DLYKQNKVLVNQLSQPEPGTSDLHFPTKYSQSTIGQF 1022
 Qy 400 TTLIRROI SNDFRDLPTLLIHGAEACLSMTIGFLYFGHGSIQLSFMDTAALLFMIGALI 459
 : : | : | : : : : : : : : : : : : : : | : | : | :
 Db 1023 RACLWKQWLTYWRSPDYNLVRFSFTLFTALLGTIFWKIGT---KMGNANS LR MVIGAM- 1078
 Qy 460 PFNVILDV-ISKYS-----ERAMLYYELEDGLYTTGPYFFAKILGELP---EHCAY- 507
 : : : : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 1079 -YTAVMFIGINNCATVQPIVSIERTVFYRERAAGMYSAMPYAIAQVMEIPYVFVQTAYY 1137
 Qy 508 -IIIIYGMPTY-WLANLRPGLQPFLHFLLVLVVFCCRIMALAAAALLPTFHMASFFSNA 565
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 1138 TLIVYAMMSFQWTA-----AKFFWFFFVSYSFLYFTYYGMMTVAISPHEVA AIFAAA 1191
 Qy 566 LYN SFYLAGGFMI---NLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNL-- 620
 | : | | | : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 1192 FYSLEFNLSRFFIPRPRIKWWIYYWLCP---LAWTVYGLIVTQY-----GDLEQ 1239
 Qy 621 TIAVSGDKILSAMELDSY-----PLYAIYLIVIGLSGGFM 655
 | : | | | : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 1240 IISVPGQ---SNQTISYYVTHHFGYHRKFMPVVPVVLVLFVAVFFAFM 1283

RESULT 5

US-09-614-912-144

; Sequence 144, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve
 ; APPLICANT: Rafalski, Antoni
 ; APPLICANT: Orozco, Buddy
 ; APPLICANT: Miao, Gou-Hau
 ; APPLICANT: Famodu, Omolayo O.
 ; APPLICANT: Lee, Jian Ming
 ; APPLICANT: Sakai, Hajime
 ; APPLICANT: Weng, Zude
 ; APPLICANT: Caimi, Perry G
 ; APPLICANT: Anderson, Shawn
 ; TITLE OF INVENTION: Plant Metabolism Genes
 ; FILE REFERENCE: BB1378 US NA
 ; CURRENT APPLICATION NUMBER: US/09/614,912
 ; CURRENT FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,401
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/143,412
 ; PRIOR FILING DATE: 1999-07-12
 ; PRIOR APPLICATION NUMBER: 60/146,650

```
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 144
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (272)..(273)
US-09-614-912-144
```

```
Query Match          12.6%; Score 440.5; DB 4; Length 539;
Best Local Similarity 25.3%; Pred. No. 8.4e-40;
Matches 147; Conservative 114; Mismatches 241; Indels 79; Gaps 18;
```

```
Qy      124 GGIKISGQIWINGQPSSPQLVRKCVAHVRQHNLNLTVRETLAFIAQMRLPRTFSQAO 183
      || |: |:| :|| | : : : :|:: |:|: |:| || :||| :
Db      1 GGYIE-GEITVSGYPKKQETFARISGYCEQNDIHSPhVTIYESLVFSAWLRLPAEVDSE 59

Qy      184 RDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPgilILDEPTSG 243
      | :|::: : | || | ||| :|:|::| |:|: || |: :| || || |
Db      60 RKMfIEEIMDLVELTSLRGALVGLPGVngLSTEQRKRLTIaVELVANPSIIFMDEPTSG 119

Qy      244 DSFTAHLNVKTLsRLAKGNRLVLISLHQPRSDIFRLFD-LVLLMTSGTPIYLGA----AQ 298
      |: | :::|: | |: :||| ||| || | |: | ||:| :
Db      120 DARAAIvmRTVRNTVNTGRTVVCTIhQPSIDIFEAFDELfLMKRGGEeIYVGPVgQNSA 179

Qy      299 HmVQYFTAI-----GYPCPRYSNPADfyVDLTsIDRRSREqELATREKAQSLAALFL 350
      ::::| | | || | ||| : ::::| | :|: | | |
Db      180 NLIeYfEEIEGISKIKDGY-----NPATWMLEvSS----SAQEEM-----LGIDFA 221

Qy      351 EKVRDLDDFLWKAETKDLDEDTcVessVtPLDTNCLPSPTKMPGA-VQQFTTLIRRQISN 409
      | | :|| ::: | |: : : | ||: : | | : :| :
Db      222 EVYR-----QSELYQRNKELIKELsMPAGSSDLNfPTQYSRSFVTQCLACLWKQXXS 274

Qy      410 DFRDLPTLLIHGAeAClMSMTIGFLYFGHGSIQLSFMDT-----AALLFM----IGA 457
      :|: : : ::: | ::: || | ||:|: : | :
Db      275 YWRNPSYTAVRLLFTIvIALMFGTMfWDLGSKTRRSQDLfNAMGSMYAAVLYIGVQNSGS 334

Qy      458 LIPFNvILDVIsKCYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIiYGMPTYW 517
      : | |: || : | | |:|: || | : : | | :|| | |
Db      335 VQPvVVV-----ERTVFYRERAAgMYSAFPYAfgQVAIEFPYVLVQALIYGGLVYS 385

Qy      518 LANLRPGLQPfLLHfLLVWLvvFCCRIMALAAAALLPTfHMASFFSNALYNSfYLAGGF 577
      : : | | : : : : | | | :|: |:| || | | :|:
Db      386 MIGFEWTVAKFLWYlFFMYFTMLYFTfyGMMaVGLTPNESIAAIISsAFYNvWNLFsGYL 445

Qy      578 I---NLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAME 634
      | | | :|| |: | ||: || : || | | :| |
Db      446 IPRPKLPIWWRWYsWICPVA---WTLYGLVASQFG--DIQQPLDQ---GVPGPQITVAQF 497
```



```
Qy      635 LDSY-----PLYAIYLIVIGLGGFMVLY-YVSLRFIKQK 668
      : |           | : : : : : | | : : : || ||
Db      498 VTDYFGFHHDFLVWVAMVHVAFTVLFQFLEFSAIMRFNFQK 538
```

Db 523 NKDRAAVRDLNLNLYEGQITVLLGHNGAGKTTTSLMLTGL---FPPTSGRAYISGYEISQ 579

QY 142 QLV--RKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQ 199
 :| || : ||: | ||| | | ||:: | | |:| | :

Db 580 DMVQIRKSLGLCPQHDILFDNLTVAEHLYFYAQLK-----GLSRQKCPREEVKQMLHIIG 633

QY 200 CADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLVKTLSRLA 259
 | | | ||| ||::|||: | : ||||| |||:|: : : | |

Db 634 LEDK--WNSRSRFLSGGMRRKLSIGIALIAGSKVLILDEPTSGMDAISRRAIWDLQR-Q 690

QY 260 KGNRLVLISLH-QPRSDIFRLFDLVLLMTS-----GTPIYL---GAAQHMVQYFTAIG 308
 | :| ::: | :|: | | : :| | : :| || || | :

Db 691 KSDRTIVLTTHFMDEADL--LGDRIAIMAKGELQCCGSSLFLKQKYGAGYHM----TLVK 744

QY 309 YPCPRYSNPADF-----YVDLTSIDRRSREQELA---TREKAQSLAALFL---EKVRDL 356
 | : || | :| ::: | ||: || || :| ::|

Db 745 EP---HCNPEDISQLVHHHVPNATLE-SSAGAELSFILPRESTHRFEGLFAKLEKKQKEL 800

QY 357 DDFLWKAETKDLDE-----DTCVESSVT-----PLDTNCLPS--P 389
 : | :| |::| :|:| :|:| : |

Db 801 GIASFGASITTMEEVFLRVGKLVDSMDIQAIQLPALQYQHERRASDWAVDSNLCGAMDP 860

QY 390 TKMPGAV-----QOFTTLIRRQISNDFRDLPTLLIHGAEC--IMSM 429
 : ||: ||| : ::: |: : | : | :

Db 861 SDGIGALIEEERTAVKLNTGLALHCQQFWAMFLKKAAYSWREWKMV---AAQVLVPLTCV 917

QY 430 TIGFLYFGHGSIQLSFMDTAALLFMIG----ALIPFNVI-LDVISKCYSERAMLYYELED 484
 |: | : | | :| :||:| : : || |:|

Db 918 TLALLAINYSS---ELFDDPMLRLTLGEYGRTVVPFSPGTSQLGQQLSE-----HLKD 968

QY 485 GLYTTG--PYFFAKILGELPE 503
 | | | :||:| |

Db 969 ALQAEQGEP---REVLGDLEE 986

RESULT 7

US-08-762-500-25

; Sequence 25, Application US/08762500

; Patent No. 6030806

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 83

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1684 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-762-500-25

```

```

Query Match          7.1%; Score 250; DB 3; Length 1684;
Best Local Similarity 24.8%; Pred. No. 1.3e-17;
Matches 124; Conservative 85; Mismatches 176; Indels 116; Gaps 27;

```

```

Qy      82 NSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSSP 141
      |   ::||:  : ||:  ::| :| | : : | ::||   ||:  ::|:|  |
Db      523 NKDRAAVRDLNLNLYEGQITVLLGHNGAGKTTTSLMLTGL---FPPTSGRAYISGYEISQ 579

Qy      142 QLV--RKCVAHVRQHNLQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQ 199
      :|  || :   ||:  |  ||||  |  |  ||::   |  |  |:|  | :
Db      580 DMVQIRKSLGLCPQHDLFDNLTVAEHLFYFQAQLK-----GLSRQKCPEEVKQMLHIIG 633

Qy      200 CADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLSRLA 259
      |   |   |  ||||  ||::|||:  |:   :||| |||||:|:  :   :  | |
Db      634 LEDK--WNSRSRFLSGGMRRKLSIGIALIAGSKVLILDEPTSGMDAISRRAIWDLLQR-Q 690

Qy      260 KGNRLVLISLH-QPRSDIFRLFDLVLLMTS-----GTPIYL----GAAQHMVQYFTAIG 308
      | :|  ::::  |   :::  | | : :|   | : ::|   ||  ||   | :
Db      691 KSDRTIVLTTFHMDL--LGDRIAIMAKGELQCCGSSLFLKQKYGAGYHM----TLVK 744

Qy      309 YPCPRYSNPADF-----YVDLTSIDRRSREQELA---TREKAQSLAALFL---EKVRDL 356
      |   : || |   :|   :::  |  ||:  ||   ||   :|  ::|
Db      745 EP---HCNPEDISQLVHHHVPNATLE-SSAGAELSFILPRESTHREFGLFAKLEKKQKEL 800

Qy      357 DDFLWKAETKDLDE-----DTCVESSVT-----PLDTNCLPS--P 389
      :  |   ::|   |::|:   :|:|  :  |
Db      801 GIASFGASITTMEEVFLRVGKLVDSMDIQAIQLPALQYQHERRASDWAVDNLCGAMDP 860

```

Qy 390 TKMPGAV-----QQFTTLIRQISNDFRDLPTLLIHGAEAC--LMSM 429
: ||: ||| : :: : |: : | :
Db 861 SDGIGALIEEERTAVKLNTGLALHCQQFWAMFLKKAAYSWREWKMV---AAQVLVPLTCV 917

Qy 430 TIGFLYFGHGSIQLSFMDTAALLFMIG----ALIPFNVI-LDVISKCYSERAMLYYELED 484
|: | : | | :| ::||: : : || |:|
Db 918 TLALLAINYSS---ELFDDPMLRLTLGEYGRTVVPFSVPGTSQLGQQLSE-----HLKD 968

Qy 485 GLYTTG--PYFFAKILGELPE 503
| | | :||:| |
Db 969 ALQAEGQEP---REVLGDLEE 986

RESULT 8

US-08-762-500-75

; Sequence 75, Application US/08762500

; Patent No. 6030806

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 83

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

; CITY: Framingham

; STATE: Massachusetts

; COUNTRY: United States of America

; ZIP: 01701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/762,500

; FILING DATE: 09-DEC-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/665,259

; FILING DATE: 17-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/10469

; FILING DATE: 17-JUN-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Dugan, Deborah A.

; REGISTRATION NUMBER: 37,315

; REFERENCE/DOCKET NUMBER: IG5-9.3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (508) 872-8400

; TELEFAX: (508) 872-5415

; INFORMATION FOR SEQ ID NO: 75:

```
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1704 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-762-500-75
```

```
Query Match          7.1%; Score 250; DB 3; Length 1704;
Best Local Similarity 24.8%; Pred. No. 1.3e-17;
Matches 124; Conservative 85; Mismatches 176; Indels 116; Gaps 27;
```

```
Qy      82 NSCELGIONLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINGQPSSP 141
      |   :::|:  : ||:  ::| :| : : | ::||           ||:  ::|:|  |
Db      543 NKDRAAVRDLNLNLYEGQITVLLGHNGAGKTTTSLMLTGL---FPPTSGRAYISGYEISQ 599

Qy     142 QLV--RKCVAHVQRHQNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQ 199
      :|  || :   ||: |  ||| |  | | ||::           | |  |:|  | :
Db      600 DMVQIRKSLGLCPQHDLFDNLTVAEHLFYFQAQLK-----GLSRQKCPEEVKQMLHIIG 653

Qy     200 CADTRVGNMVVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLSRLA 259
      |   |   |  ||| |  ||: |||: | :   : ||| ||| ||: | :   : | |
Db      654 LEDK--WNSRSRFLSGGMRRKLSIGIALIAGSKVLILDEPTSGMDAISRRAIWDLQR-Q 710

Qy     260 KGNRLVLISLH-QPRSDIFRLFDLVLLMTS-----GTPIYL----GAAQHMVQYFTAIG 308
      | :|  :::: |   ::: | | : :|           |: ::|   ||  ||   | :
Db      711 KSDRTIVLTTHFMDEADL--LGDRIAIMAKGELQCCGSSLFLKQKYGAGYHM----TLVK 764

Qy     309 YPCPRYSNPADF-----YVDLTSIDRRSREQELA---TREKAQSLAALFL---EKVRDL 356
      |   : || |   :|   ::: |   ||:   ||           ||   :|  ::|
Db      765 EP---HCNPEDISQLVHHHVPNATLE-SSAGAELSFILPRESTHREGLFAKLEKKQKEL 820

Qy     357 DDFLWKAETKDLDE-----DTCVESSVT-----PLDTNCLPS--P 389
      : |   ::|           |::|:           :|:|   : |
Db      821 GIASFGASITTMEEVFLRVGKLVDSMDIQAIQLPALQYQHERRASDWAVDSNLCGAMDP 880

Qy     390 TKMPGAV-----QFFTTLIRRQISNDFRDLPTLLIHGAEC--LMSM 429
      :  ||:           ||| :   : :|:   : |   | :
Db      881 SDGIGALIEEERTAVKLNTGLALHCQQFWAMFLKKAAYSWREWKMV---AAQVLVPLTCV 937

Qy     430 TIGFLYFGHGSIQLSFMDTAALLFMIG----ALIPFNVI-LDVISKCYSERAMLYYELED 484
      |: |   : |   |   :|   ::||:|   : :  ||           |:|
Db      938 TLALLAINYSS---ELFDDPMLRLTLGEYGRTVVPFSVPGTSQLGQQLSE-----HLKD 988

Qy     485 GLYTTG--PYFFAKILGELPE 503
      |   |   |   ::||:| |
Db      989 ALQAEGQEP---REVLGDLEE 1006
```

RESULT 9

US-09-489-039A-10393

; Sequence 10393, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

```
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10393
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10393
```

```
Query Match          6.9%; Score 243.5; DB 4; Length 265;
Best Local Similarity 28.8%; Pred. No. 2.6e-18;
Matches 66; Conservative 49; Mismatches 97; Indels 17; Gaps 4;
```

```
Qy      86 LGIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINGQPSSPQLVR 145
      | :||:| : | : ::|| |||:::|:| : | : | : : | |
Db      25 LALQNVSF DIVEGETISLIGHSGCGKSTLLNLIA--GITTPTEGGLLCDNREIAGPGPER 82

Qy     146 KCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRV 205
      | | |: : ||| | : : | | | : : | : | : : | |
Db      83 ---AVVFQNHSLLPWLSCFDNVALAVDQVFRRTMSKSERREWIEHNLARVQMGMHALHKRP 139

Qy     206 GNMVYVRGLSGGERRRVSIGVQLLWNP GILILDEPTSGLD SFTAHNLVKTL SRLAKGNRLV 265
      |      :||| :|| | | | | :|:| || | | : | | : :
Db     140 GE-----ISGGMKQRVGIARALAMKPKVLLLDEPFGALDALTRAHLQDTVMHIQQELNTT 194

Qy     266 LISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314
      :: : : | | ||:|:| : | | : | | | :
Db     195 IVMITHDVDEAVLLSDRVLLMTNGPAATVGE-----ILAVDLPRPRH 236
```

RESULT 10

```
US-09-252-991A-21665
; Sequence 21665, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21665
; LENGTH: 593
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21665
```

```
Query Match          6.9%; Score 243; DB 4; Length 593;
```

Best Local Similarity 27.5%; Pred. No. 1.2e-17;
Matches 74; Conservative 61; Mismatches 100; Indels 34; Gaps 9;

```
Qy      89 QNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQ---PSSPQLVR 145
      : : | | | : : : | | | : : | | | : | | :
Db      282 RDIDFAAARGEFTLLGSPGCGKSTLLRCIAGL---TEVDSGRILIDGEDVVPLPPQ--K 336

Qy      146 KCVAHVRQHNQLLPNLTVRETLAF-IAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTR 204
      : : | | | | | : : : : : | : : | : | : | : |
Db      337 RGIAMVFQSYALFPNMTVQQNVAFLRMQKVP----AAELKQRVAEAIELVELGEYA--- 389

Qy      205 VGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLSRLAKGNRL 264
      | | | : : | : | : | : | | | : | : : | : : |
Db      390 --ARYPHQLSGGQCQRVALARSLVTRPRLLLDEPLSALDARIRKHLREQIRRIQQELGL 447

Qy      265 VLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFY--V 322
      : : : : | | : : | : | : | : : | | : | |
Db      448 TTVFVTHDQEEALTLSDRIVLMNAGRIVQSGDAETL-----YTAPENAFAGFIGNY 499

Qy      323 DLTSIDRRSR-----EQELATREKAQSL 345
      : | : : | | | : : | : : |
Db      500 NLLDAEQASRLLGQPCAQQVAIRPESLRL 528
```

RESULT 11

US-09-252-991A-27569

; Sequence 27569, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 27569

; LENGTH: 330

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27569

Query Match 6.9%; Score 242; DB 4; Length 330;

Best Local Similarity 27.5%; Pred. No. 5.7e-18;

Matches 85; Conservative 60; Mismatches 120; Indels 44; Gaps 11;

```
Qy      88 IQNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQQPSSPQLVR-K 146
      : | : : : | : : | | | : : | : | : | : | : |
Db      19 LDNINLDIQSGELVALLGSPGCGKTTLLRIIAGL---ETPDAGNIVFHGEDVSQHDVRDR 75

Qy      147 CVAHVRQHNQLLPNLTVRETLAFIAQMRLPR--TFSQAQRDKRVEDVIAELRLRQCADTR 204
      | | | | | : : | : : | : : | : : : | : : : | |
```

```

Db      76 NVGFVFQHYALFRHMTVFDNVAFLRMK-PKGERPGESAIKAKVHELLNMVQLDWLAD-- 132
Qy      205 VGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLNVLKTLRLAKGNRL 264
      | | | | : | : : : | | | : | | | | : | : | : | : |
Db      133 ---RYPEQLSGGQRQRIALARALAVEPKILLLDEFFGALDAKVRKELRRWLARLHEEINL 189
Qy      265 VLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPA-DF-YV 322
      : : : : : | : : : | | | | | | | | | | | | | |
Db      190 TSVFVTHDQEEAMEVADRIVVMNKGV-----IEQIGSPGEVYENPASDFVYH 236
Qy      323 DLTSIDR-----RSREQELATREKAQSLAALFLEKVRDLDDF--LWKAETKDLD 369
      | : | | | | | : | : | | : | | : | | : |
Db      237 FLGDSNRLQLGNDQHLLFRPHEVSLSRSEVAEHRAA----EVRDIRPLGAITRVTLKVDG 292
Qy      370 EDTCRESSV 378
      : | : | : |
Db      293 QDELIEAEV 301

```

RESULT 12

```

US-09-489-039A-11991
; Sequence 11991, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11991
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11991

```

```

Query Match          6.8%; Score 238; DB 4; Length 379;
Best Local Similarity 28.0%; Pred. No. 2e-17;
Matches 69; Conservative 48; Mismatches 89; Indels 40; Gaps 9;

```

```

Qy      88 IQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRG--HGGKIKSGQIHING-QPSSPQLV 144
      : : | : | : : | : | | | : : | : | | : : | : |
Db      24 VHGIDLKIADGEFMVIVGPSGCAKSTTLRMLAGLETISGGEVRIGDKIVNNLAPKS---- 79
Qy      145 RKCVAHVRQHNQLLPNLTVRETAF-IAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADT 203
      : : | | | : | | : | | | | | : : | : | | : | : |
Db      80 -RGIAMVFQNYALYPHMTVRENLAFLGLKLSKLPK----AQIDRQVEEAAKILELEELLD- 133
Qy      204 RVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLNVLKTLRLAKGNR 263
      | | | | : | : : : | : : | : : | | | | | : : |
Db      134 ----RLPRQLSGGQAQRVAVGRAIVKKPDVFLFDEPLSNLD-----AKLRASMR 178
Qy      264 LVLISLH-----QPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSN 316

```



```

      : : ||      :| : :: |      || | | :      |::| |      |      |
Db      179 IRISDLHKQLKKSQKGPATTVYVTHDQTEAMTMGDRICVMKLGHHIMQVDT----PDNLYHQ 234

Qy      317 PADFYV 322
      | : :|
Db      235 PKNMFV 240

```

RESULT 13

```

US-09-489-039A-8815
; Sequence 8815, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8815
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8815

```

```

Query Match          6.6%; Score 231; DB 4; Length 388;
Best Local Similarity 27.1%; Pred. No. 1.3e-16;
Matches 69; Conservative 53; Mismatches 87; Indels 46; Gaps 9;

```

```

Qy      86 LGIQN-----LSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQI 132
      | :||          || : | : : :| ||| : : || : : |      : |||
Db      34 LSLQNISKRFDGKPAALSALSLDIHEGEFVVLVGPSGCGKSTLLRLLAGL---EPVSEGQI 90

Qy      133 WINGQ---PSSPQLVRKCVAVHRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVE 189
      |:: :      ::| : : | : | : | :| :| : : | : :|      ::      ||:
Db      91 WLHNENITAATPR--ERNFAMIFQNYALFPHLSVRDNITFGMKVRKE---EKSSWQPRVD 145

Qy      190 DVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTFAH 249
      |      | :|      | :      ||| :| :| : : : : || : : ||| | ||
Db      146 KVAQMLQLEALLDRKPAK-----LSGGQRQVRAMARAIVRNPRFLMDEPLSNLD----- 195

Qy      250 NLVKTL SRLAKGNRLVLISLHQ--PRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAI 307
      :||      | : :|||      | | : |      | :      | : |      :
Db      196 -----ARLRSEVRDSIMALHQQLKTSTIYVTHDQTEAMSMADRIVVMNGGHVQQ-----V 245

Qy      308 GYPCPRYSNPADFYV 322
      | |      | :||| : :|
Db      246 GRPEYLYANPANLFV 260

```

RESULT 14

```

US-09-252-991A-20719
; Sequence 20719, Application US/09252991A

```

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20719
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20719
```

```
Query Match          6.6%; Score 230; DB 4; Length 370;
Best Local Similarity 27.4%; Pred. No. 1.5e-16;
Matches 83; Conservative 55; Mismatches 121; Indels 44; Gaps 10;
```

```
Qy      88 IQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPSQPQLVRKC 147
      :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db      28 VDNVSLTINTGEFFTLGLPSGCGKTTLLRMLAG---FDQPDSGEIRLNGQDLAGEVEPEKR 84

Qy     148 VAH-VRQHNQLLPNLTVRETIAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVG 206
      ||| | | |::| :|:| :| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db      85 PVHTVFQSYALFPHMSVAQNIAPFLKM---AGVAKSEIDARVEQALKDVRL---ADK--G 136

Qy     207 NMYVRGLSGGERRRVSIGVQLLWNPGLILILDEPTSGLDSTAHNLVKTLRLAKGNRLVL 266
      |||:|:|:| | :| :|:|:| | | :| :| :| :| :| :| :| :| :| :| :|
Db     137 GRMPTQLSGGQRQORVAIARALVNRPRLLLLDEPLSALDAKLREEMQIELINLQKDVGITF 196

Qy     267 ISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADF-----Y 321
      : : : : | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     197 VYVTHDQGEALALSHRIAVMNQGRVEQLDAPETI-----YSFPRSRFVADFIGQCNL 248

Qy     322 VDLT--SIDRRSREQELATREKAQSLAA-----LFLEKVRDLDDFLWKAETKD 367
      :| | :|:| :| :| :|:| :| :|:| :| :|:| :| :|:| :| :|:| :| :|
Db     249 LDATVEAVDGERVRIDLRLGLGEVQALKSFDAQPGACVLTTLRPEKIR-----LAQSVTAD 303

Qy     368 LDE 370
      ||
Db     304 SDE 306
```

RESULT 15

```
US-09-134-000C-3584
; Sequence 3584, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
```

```
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3584
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-3584
```

```
Query Match          6.5%; Score 229; DB 4; Length 229;
Best Local Similarity 27.8%; Pred. No. 8.5e-17;
Matches 70; Conservative 56; Mismatches 88; Indels 38; Gaps 8;
```

```
Qy      47 LEVRDLNYQVDLASQVPWFEEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSGQMLAIIGS 106
        |||||:          : : | :      :: : :: ||| |: |: :|:|
Db      3 LEVRDM-----ANVLEMKNIIKKYGEKHTEVIALKELSFVQPGFVAVIGP 49

Qy     107 SGCGRASLLDVITGRGHGGKIKSGQIWINQPSQPQLVRK-----CVAHVRQHNQLL 158
        || |::: | : |      ||:: : || | :| :|      : : | : |:
Db     50 SGSGKSTFLTIAAGL---QAPTSGEVIVGGQ-SLNKLTKKQRLAQRFQKIGFILQSSNLV 105

Qy     159 PNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGER 218
        | ||| :      | :: | | : :      | :: | |::      : | | | |||||
Db     106 PFLTVEDQFHLIEKVDKSRKNSELK-----EQLETLGLKE-----LRNSYPRDLSGGER 155

Qy     219 RRVSIGVQLLWNPGILILDEPTSGLDSTAHNLVKTLRLAK-GNRLVLISLHQPRSDIF 277
        :||| |      | ::: |||: ||: | :|| |:: ||      :: ::: | | :
Db     156 QRVSIACALYHEPDVILADEPTASLDTEKAFDVVKLLAKEAKEKDKGIIMVTHDER--LL 213

Qy     278 RLFDLVLLMTSG 289
        : | | : |
Db     214 KYCDRVVRIRDG 225
```

```
Search completed: February 27, 2004, 07:20:17
Job time : 16.2492 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:44:33 ; Search time 14.9951 Seconds
(without alignments)
4317.206 Million cell updates/sec

Title: US-09-989-981A-8
Perfect score: 3506
Sequence: 1 MAGKAAEERGLPKGATPQDT.....FMVLYYVSLRFIKQKPSQDW 673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	735.5	21.0	725	2	C84423	probable ABC trans
2	723.5	20.6	646	2	C86441	probable ABC trans
3	700	20.0	609	2	E96742	probable ABC trans
4	668.5	19.1	737	2	T46101	ABC transporter-li
5	656	18.7	687	1	FYFFW	white protein - fr
6	653	18.6	656	2	JC7860	brain multidrug re
7	653	18.6	1294	2	S77690	probable membrane
8	651.5	18.6	635	2	T08934	hypothetical prote
9	618	17.6	638	2	G02068	white homolog - hu
10	595	17.0	687	2	D96553	hypothetical prote
11	591	16.9	725	2	T47652	ABC transporter-li
12	590	16.8	559	2	B88474	protein C05D10.3 [
13	589.5	16.8	720	2	T47648	ABC transporter-li

14	586.5	16.7	708	2	T47650	ABC transporter-li
15	579	16.5	658	2	T31958	hypothetical prote
16	575.5	16.4	678	2	H96552	hypothetical prote
17	573.5	16.4	646	2	JC7777	ATP binding casset
18	562.5	16.0	1049	1	S19421	ATP-dependent perm
19	561.5	16.0	739	2	T45891	ABC transporter-li
20	561.5	16.0	755	2	G84791	probable ABC trans
21	556.5	15.9	740	1	T02567	probable ATP-bindi
22	550	15.7	649	2	A84509	probable ABC trans
23	548	15.6	662	2	T47649	ABC transporter-li
24	547	15.6	590	2	B96573	protein F12M16.17
25	539	15.4	695	2	T21109	hypothetical prote
26	534	15.2	633	2	T19189	hypothetical prote
27	532.5	15.2	659	2	E86313	hypothetical prote
28	527	15.0	1443	2	T02491	probable ABC trans
29	525.5	15.0	577	2	T04229	ABC-type transport
30	524	14.9	608	2	T34391	hypothetical prote
31	521	14.9	639	2	G88839	protein C10C6.5 [i
32	511.5	14.6	547	2	T31543	hypothetical prote
33	511	14.6	610	2	T19333	hypothetical prote
34	506.5	14.4	1423	2	A86289	probable ABC trans
35	480	13.7	1450	2	T45888	ABC transporter-li
36	474	13.5	1451	2	B86286	F9L1.15 protein -
37	467.5	13.3	1420	2	T02644	ABC-type transport
38	466.5	13.3	1426	2	T30567	ATP-binding casset
39	464	13.2	1564	2	S55517	probable transport
40	463.5	13.2	675	1	FYFFB	brown protein - fr
41	461	13.1	1469	2	H96622	probable ABC trans
42	452	12.9	1435	2	D96693	protein Putative A
43	446	12.7	1450	2	A84780	probable ABC trans
44	440.5	12.6	1413	2	G84790	probable ABC trans
45	437	12.5	668	2	S55023	brown protein - fr

ALIGNMENTS

RESULT 1

C84423

probable ABC transporter [imported] - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: C84423

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84423

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-725 <STO>

A;Cross-references: GB:AE002093; NID:g4262239; PIDN:AAD14532.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g01320

A;Map position: 2

Query Match 21.0%; Score 735.5; DB 2; Length 725;
Best Local Similarity 30.0%; Pred. No. 3.6e-50;
Matches 186; Conservative 123; Mismatches 229; Indels 81; Gaps 15;

```
Qy      75 WTSPSC-----QNSCELGIONLSFKVRSGQMLAIIGSSGCGRASLLDVTG-----RG 122
      | : :|      |      ::|| : : ||:||||| || | : :||:| : |
Db      72 WRNITCSLSDKSSKSVRFLLNKVSGEAKPGRLLAIMGPSGSGKTTLLNVLAGQLSLSPRL 131

Qy     123 HGGKIKSGQIWINGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQA 182
      |      || : :||:|||| : : :| ||| :      |||||:| | :||| |
Db     132 H----LSGLLEVNGKPSSSKAYK--LAFVRQEDLFFSQLTVRETLSFAAELQPEISSAE 185

Qy     183 QRDKRVEDVIAELRLRCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSG 242
      :||: | :||: | | |||: ||: |||:||||:|:|: :|: :| :| :| |||:|
Db     186 ERDEYVNNLLLKLGLVSCADSCVGDAKVRGISGGEKKRLSLACELIASPSVIFADEPTTG 245

Qy     243 LDSFTAHNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLG-AAQHMV 301
      ||:| | :||| :||: | : ||||| :| || :||:| || :| | | :
Db     246 LDAFQAEKVMETLQKLAQDGHTVICSIHQPRGSVYAKFDDIVLLTEGTLVYAGPAGKEPL 305

Qy     302 QYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFLW 361
      || | : || : |||:| || | :| | | :||: :| | :|
Db     306 TYFGNFGFLCPEHVNPAEFLADLISVDYSSSETVYSSQKRVALVDAFSQR----- 356

Qy     362 KAETKDLDEDTCEVSSV---TPLDTNCLPSPTK-----MPGAVQQFTTLIRR 405
      ||| ||| : ||      | :|| | :|
Db     357 -----SSSVLYATPLS---MKEETKNGMRPRRKAIVERTDGWWRQFFLLLKR 400

Qy     406 QISNDFRDLPTLLIHGAACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVIL 465
      || || : : :| :|| | | | | || : | : | :
Db     401 AWMQASRDGPTNKVRARMSVASAVIFGSVFWRMGKSQTSIQDRMGLL-QVAAI---NTAM 456

Qy     466 DVISKCY----SERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANL 521
      :||      |||:| | | :| || :| : | :| | : :|| | :| |
Db     457 AALTKTVGVFPKERAIVDRERSKGSYSLGPYLLSKTIAEIPIGAAPPLMFGAVLYPMARL 516

Qy     522 RPGLQPFLHFLLVVLVVFCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLS 581
      | | | :| : | | | | :||: | :| | :| :| :| :|
Db     517 NPTLSRFGKFCGIVTVESFAASAMGLTVGAMVPSTEAMAVGPSLMTVFIVFGGYVNAD 576

Qy     582 SLWTVPAWISKVSFLRWCFEGLMKIQFS-----RRTYKMPLGNLT---IAVSGDKILSA 632
      : : || : | :|| | :|| :|| :| : : | :| :| :|
Db     577 NTPIIFRWIPRASLIRWAFQGLCINEFSGLKFDHQNTFDVQTGEQALERLSFGGRRIRET 636

Qy     633 MELDSYPLY----AIYLIV 647
      : | | | ||:|
Db     637 IAAQSRIIMFWYSATYLLL 655
```

RESULT 2
C86441

probable ABC transporter [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C;Accession: C86441
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;
 Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;
 Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;
 Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;
 Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;
 Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;
 Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-
 Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,
 S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,
 M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;
 Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,
 S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;
 Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A;Reference number: A86141; MUID:21016719; PMID:11130712
 A;Accession: C86441
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-646 <STO>
 A;Cross-references: GB:AE005172; NID:g11136734; PIDN:AAG31315.1; GSPDB:GN00141
 C;Genetics:
 A;Map position: 1
 C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein
 F12L6.1; ATP-binding cassette homology

Query Match 20.6%; Score 723.5; DB 2; Length 646;
 Best Local Similarity 30.6%; Pred. No. 2.7e-49;
 Matches 208; Conservative 119; Mismatches 262; Indels 91; Gaps 20;

Qy	22	GLQDRLEFSSES DNSLYF-TYSGQPN-----TLEVRDLNYQVDLASQVPWF EQLAQFKMP	74
		: : : : : : : : : : :	
Db	20	GLPD-MSDTQSKSVLAFPTITSQPGLQMSMPITLKEVVYKVKI-----EQTSQCMGS	71
Qy	75	WTSPSCQNSCELGIQNL SFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWI	134
		: : : : : : : : : :	
Db	72	WKSKE-----KTILNGITGMVCPGEFLAMLGPSGSGKTTLLSALGGR--LSKTFSGKVMY	124
Qy	135	NGQPSSPQLVRKCVAHVRQHNQLLPNLT VRET LAFIAQMRLPRTFSQAQRDKRVEDVIAE	194
		: : : : : : : : : :	
Db	125	NGQPFSGCIKRR-TGFVAQDDVLYPHLTVWETLFFTALLRLPSSLTRDEKAEHVDRVIAE	183
Qy	195	LRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPTSGLD SFTAHNLVKT	254
		: : : : : : : :	
Db	184	LGLNRCTNSMIGGPLFRGISGGEKKRVSIGQEMLINPSLLLLDEPTSGLDSTTAHRIVTT	243
Qy	255	LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY	314
		: : : : : : : : : : : : : :	
Db	244	IKRLASGGRTVVTTIHQPSSRIYHMFDKVLLSEGSPIIYGAASSAVEYFSSSLGFSTSLT	303

A;Residues: 1-609 <STO>
A;Cross-references: GB:AE005173; NID:g6978921; PIDN:AAF34313.1; GSPDB:GN00141
C;Genetics:
A;Gene: F17M19.11
A;Map position: 1
C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 20.0%; Score 700; DB 2; Length 609;
Best Local Similarity 31.8%; Pred. No. 1.8e-47;
Matches 210; Conservative 103; Mismatches 225; Indels 122; Gaps 22;

```

Qy      76 TSPSCQNSCELGI-QNLSFKVRS-----GQMLAIIGSSGCGRASLLDVI 118
      :: ||      ||: | | : ||      |: :|:| || |::||: :
Db      2  SNDSCNIKKLLGLKQKPSDETRSTEERTILSGVTGMISPGEFMAVLGSPSGSGKSTLLNAV 61

Qy     119 TGRGHGGKIKSGQIWINQPPSPQLVRKCAVHVRQHNQLLPNLTVRETAFIAQMRLPRT 178
      || || : :|:| || : |:: : | | : | |:| ||||| |:| :|||:
Db     62  AGRLHGSNL-TGKILINDGKITKQTLKR-TGFVAQDDLLYPHLLTVRETIVFVALLRLPRS 119

Qy     179 FSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDE 238
      :: : : | ||:| | :| :| ||| ::|:|:| |||:| || :| || :|:|
Db     120 LTRDVKLRAAESVISELGLTKCENTVVGNTFIRGISGGERKRVSIHELLINPSLLVLDE 179

Qy     239 PTSGLDSFTAHLNVLKTLRLAKG-NRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGA 297
      |||||: | ||:|: || | : | :|:| | :|:| |||:| :|:|
Db     180 PTSGLDATAALRLVQTLAGLAHGKGTVVTSIHQPSSRVFQMFDTVLLLSEKCLFVGKG 239

Qy     298 QHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREK---AQSLAALFLEKVR 354
      : : || :|: | |||| :|| : : : ||| |:| :
Db     240 RDAMAYFESVGFSPAFFMPNPADFLDLA--NGVCQTDGVTEREKPNVRQTLVTAY----- 292

Qy     355 DLDDFLWKAETKDLDEDTCESSVTPLDTNCLPSPKMPGAVQQFTTLI----- 403
      | | : | ||:| | | | | | :| | | |
Db     293 ---DTLLAPQVK-----TCIEVSHFPQD-NARFVKTRVNGG--GITTCIATWFSQLCILL 341

Qy     404 -----RRQISNDFRDLPTLLIHGAEACLSMTIGFLYFGHGSIQLSFMDTAALLFMI-- 455
      || | | | : : : | : | :| : | : | | |
Db     342 HRLKERRHESFD-----LLRIFQVVAASILCGLMWW-HSDYR-DVHDRLGLLFFISI 392

Qy     456 --GALIPFNVILDVISKCYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGM 513
      | | || : |||: | | :|| || | :|| | :
Db     393 FWGVLPSEFNAVFTF---PQERAI FTRERASGMYTLSSYFMAHVLGSLSMELVLPASFLT 448

Qy     514 PTYWLANLRPGLQPFLHLLVWLVVFCCRIMALAAAALLPTFHMASFFSNALYNSFYLA 573
      |||: |||: ||| :| | : : || | : || :| |
Db     449 FTYWMVYLRPGIVPFLTLVLLLYVLASQGLGLALGAAIMDAKKASTIVTVMALAFVLT 508

Qy     574 GGEMINLSSLWTVPA---WISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVSGDKIL 630
      ||: :| ||: |: || :|: |: ||: ||:|
Db     509 GGYVYN---KVP SGMVWMKYVSTTFYCYRLLVAIQG-----SGEIL 548

Qy     631 SAMELDSYPLYA-----IYLIVIGLSG-----GFMVLYYVSLRFIK 666
      : || : ||| | | : || | :| |
Db     549 RMLGCD SKGKGASAATSAGCRFVEEEVIGDVGMWTSVGVFLMFFGYRVLAYLALRRIK 608

```

RESULT 4

T46101

ABC transporter-like protein - Arabidopsis thaliana

N;Alternate names: protein T25B15.80

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C;Accession: T46101

R;Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.

submitted to the Protein Sequence Database, January 2000

A;Reference number: Z23021

A;Accession: T46101

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-737 <ALC>

A;Cross-references: EMBL:AL132972

A;Experimental source: cultivar Columbia; BAC clone T25B15

C;Genetics:

A;Map position: 3

A;Introns: 122/1; 146/3; 225/2; 277/2; 338/3; 422/2; 535/1; 628/3; 664/3

A;Note: T25B15.80

Query Match 19.1%; Score 668.5; DB 2; Length 737;

Best Local Similarity 28.0%; Pred. No. 7.7e-45;

Matches 189; Conservative 134; Mismatches 258; Indels 93; Gaps 16;

```
Qy      25 DRLFSSSDNSLYFTYSGQPN---TLEVRDLNYQVDLASQVPWFQQLAQFKMPWTSPSCQ 81
      | | | : | : : | | : | : | | : |
Db      120 DILEDIEAATSSVVKFQAEPTFPIYLKFIDITYKV-----TTKGMT 160

Qy      82 NSCELGIQN-LSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPS 140
      : | | | : | | : | : | : | | | : | : | : | : |
Db      161 SSSEKSIILNGISGSAYPGELLALMGPSGSGKTTLLNALGGRFNQQNI-GGSVSYNDKPYS 219

Qy      141 PQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQC 200
      | : : | | : | | : | : | : | : | : | : | : |
Db      220 KHLKTR-IGFVTQDDVLFPHLTVKETLTYTALLRLPKTLTEQEKEQRAASVIQELGLERC 278

Qy      201 ADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLSRLAK 260
      | | : | : | : | : | : | : | : | : | : | : |
Db      279 QDTMIGGSFVRGVSGGERKRVICIGNEIMTNPSLLLLDEPTSSLDSTTALKIVQMLHCIAK 338

Qy      261 GNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPAF 320
      : : : : | | : | | : : : : | : : | | : | : | : |
Db      339 AGKTIVTTIHQPSSRLFHRFDKLVVLSRGSLLYFGKASEAMSYFSSIGCSPLLAMNPAEF 398

Qy      321 YVDLTSIDRRSREQELATREKAQSL-AALFLEKVRDLDDFLWKAETKDLDEDTCESSVT 379
      : | | : : | : | : : | : : | : | : | : | : :
Db      399 LLDLVNGNMNDISVPSALKEKMKIIRLELYVRNVK-----CDVETQYLEE--AYKTQIA 450

Qy      380 PLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGAACLMSM----- 429
      : : | : | : | : | : | : | : | : | : | : |
Db      451 VMEKMKLMAVPPLDEEVKLIMITCPKREWG-----LSWWEQYCLLSLRGIKERRHDY 501

Qy      430 -----TIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSER 475
      : | : : | : | : | : | : | : | : | : |
Db      502 FSWLRVTQVLSTAILGLLWW-QSDITSQRPTRSGLLFFIAVFWGFFPVFTAIFTFPQER 560
```

Qy 476 AMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLV 535
 ||| | | :| || |: :|| :: : |::| || : | | | |
 Db 561 AMLSKERESNMYRLSAYFVARTTSDLPLDLILPVLFLVVVYFMAGLRLRAESFFLSVLTV 620

Qy 536 WLNVFCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVP---AWISK 592
 :| : : : || | | |: :: :| |||: : || |||
 Db 621 FLCIVAAQGLGLAIGASLMDLKKATTLASVTVMTFMLAGGYFVK-----KVPFFIAWIRF 675

Qy 593 VSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSG 652
 :|| :: |:|:| : : |:|:| | :: : |: ::|
 Db 676 MSFNYHTYKLLVKVQYE-----EIMESVNGEEIESGLK----EVSALVAMII---- 718

Qy 653 GFMVLYYVSLRFIK 666
 |: :: | ||| :|
 Db 719 GYRLVAYFSLRRMK 732

RESULT 5

FYFFW

white protein - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 31-Dec-1990 #sequence_revision 17-Feb-1995 #text_change 19-Jan-2001

C;Accession: S08635; S07263; S10240

R;Pepling, M.; Mount, S.M.

Nucleic Acids Res. 18, 1633, 1990

A;Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.

A;Reference number: S08635; MUID:90221897; PMID:2109311

A;Accession: S08635

A;Molecule type: mRNA

A;Residues: 1-687 <PEP>

A;Cross-references: EMBL:X51749; NID:g8825; PIDN:CAA36038.1; PID:g8826

R;O'Hare, K.; Murphy, C.; Levis, R.; Rubin, G.M.

J. Mol. Biol. 180, 437-455, 1984

A;Title: DNA sequence of the white locus of *Drosophila melanogaster*.

A;Reference number: S07263; MUID:85134865; PMID:6084717

A;Accession: S07263

A;Molecule type: DNA

A;Residues: 1-24, 'LIFEIPYHCRVTAD', 30-

334, 'ITLHLNSYPAWVPSVLPTTIRRTFTYRCWPLCPDGRSSPVIGSPRYG', 372-687 <OHA1>

A;Cross-references: EMBL:X02974

A;Experimental source: strain Canton S

R;O'Hare, K.

submitted to the EMBL Data Library, June 1985

A;Reference number: S10240

A;Accession: S10240

A;Molecule type: DNA

A;Residues: 1-24, 'LIFEIPYHCRVTAD', 30-687 <OHA2>

A;Cross-references: EMBL:X02974; NID:g10873; PIDN:CAA26716.1; PID:g10874

A;Experimental source: strain Canton S

C;Genetics:

A;Gene: white; w

A;Cross-references: FlyBase:FBgn0003996

A;Introns: 24/3; 116/1; 334/2; 439/3; 483/3

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

C;Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein

F;113-317/Domain: ATP-binding cassette homology <ABC>

F;130-137/Region: nucleotide-binding motif A (P-loop)

Db 501 KPAVGSFFIMMFTLMMVAYSASSMALAIAAGQSVSVATLLMTISFVFMIFSGLLVNLK 560
 QY 582 SLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT-----IAVSGDKIL--SA 632
 :: :|: | |: | | :| : : | |:| :| : |
 Db 561 TVVPWLSWLQYFSIPRYGFSALQYNEFLGQNF-CPGLNVTNNTCSFAICTGAEYLENQG 619
 QY 633 MELDSYPLYAIYLIVIGLSGGFMVLYYVSLRFIKQ 667
 : | :: |: :: : : |: : |: | ::
 Db 620 ISLSAWGLWQNHVALACMMVIFLTIAYLKLLLLKK 654

RESULT 7

S77690

probable membrane protein YOL075c - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein O1125; hypothetical protein O1130;

hypothetical protein YOL074c

C;Species: *Saccharomyces cerevisiae*

C;Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Apr-2002

C;Accession: S77690; S66767; S66768

R;Alexandraki, D.; Katsoulou, C.; Tzermia, M.

submitted to the Protein Sequence Database, July 1996

A;Reference number: S66756

A;Accession: S77690

A;Molecule type: DNA

A;Residues: 1-1294 <ALE>

A;Cross-references: EMBL:Z74816; MIPS:YOL075c

A;Note: this is a revision to the sequence from reference S66756

A;Accession: S66767

A;Molecule type: DNA

A;Residues: 1-179, 'TTRTG VFLVVKRED' <ALW>

A;Cross-references: EMBL:Z74816

A;Experimental source: strain S288C

A;Note: this sequence has been revised in reference S77690

A;Note: this was assumed to be protein YOL074c

A;Accession: S66768

A;Molecule type: DNA

A;Residues: 200-1294 <ALF>

A;Cross-references: EMBL:Z74817

A;Experimental source: strain S288C

A;Note: this sequence has been revised in reference S77690

A;Note: this was assumed to be the complete sequence of protein YOL075c

C;Genetics:

A;Cross-references: SGD:S0005435

A;Map position: 15L

A;Note: YOL075c

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C;Keywords: ATP; nucleotide binding; P-loop; transmembrane protein

F;45-263/Domain: ATP-binding cassette homology <ABC1>

F;62-69/Region: nucleotide-binding motif A (P-loop)

F;376-392/Domain: transmembrane #status predicted <TM1>

F;469-485/Domain: transmembrane #status predicted <TM2>

F;496-512/Domain: transmembrane #status predicted <TM3>

F;606-622/Domain: transmembrane #status predicted <TM4>

F;710-916/Domain: ATP-binding cassette homology <ABC2>

F;727-734/Region: nucleotide-binding motif A (P-loop)

F;1042-1058/Domain: transmembrane #status predicted <TM5>

F;1125-1141/Domain: transmembrane #status predicted <TM6>
 F;1177-1193/Domain: transmembrane #status predicted <TM7>
 F;1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match 18.6%; Score 653; DB 2; Length 1294;
 Best Local Similarity 30.1%; Pred. No. 2.9e-43;
 Matches 171; Conservative 111; Mismatches 239; Indels 48; Gaps 13;

Qy	88 IQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQI-----	132
	: : : : : : : :	
Db	45 VNTFSMDLPSGSVMAMVGGSGGKTTLLNVLASKISGGLTHNGSIRYVLEDTGSEPNETE	104
Qy	133 ----WINGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKR-	187
	: : : : : : : :	
Db	105 PKRAHLDGQ-DHPIQKHVIMAYLPQQDVLSPRLTCRETLKFAADLKL----NSSERTKKL	159
Qy	188 -VEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSF	246
	: : : : : : : : :	
Db	160 MVEQLIEELGLKDCADTLVGDNSHRGLSGGEKRRLSIGTQMISNPSIMFLDEPTTGLDAY	219
Qy	247 TAHNLVKTL SRLAK-GNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFT	305
	: : : : : : : : :	
Db	220 SAFLVIKTLKKLAKEDGRTFIMSIHQPRSDILFLDQVCILSKGNVVYCDKMDNTIPYFE	279
Qy	306 AIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFLWKAET	365
	: : : : : : : : : : :	
Db	280 SIGYHVPQLVNPADYFIDLSSVDSRSDKEEAATQSRLNSL----IDHWHD-----YERTH	330
Qy	366 KDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGAEC	425
	: : : : : : : : :	
Db	331 LQLQAESYI-SNATEIQIQNM--TTRLP-FWKQVTVLTRRNFKLNFSYVTLISTFAEPL	386
Qy	426 LMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIP--FNVILDVISKCYSERAMLYYELE	483
	:: :: : : : : : : : :	
Db	387 IIGTVCGWIYYKPKDKSSIGGLRTTTACLYASTILQCYLYLLFDTYRLCEQDIALYDRERA	446
Qy	484 DGLYTTGPYFFA-KILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWLWVFCC	542
	: : : : : : : : :	
Db	447 EGSVTPLAFIVARKISLFLSDDFAMTMIFVSITYFMFGLEADARKFFYQFAVVFLCQLSC	506
Qy	543 RIMALAAAALLPTFHMAFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFEG	602
	::: : : : : : : :	
Db	507 SGLSMLSVAVSRDFSKASLVGNMTFTVLSMCGFFVNAKMPVYVRWIKYIAFTWYSFGT	566
Qy	603 LMKIQFSRR-----TYKMPLGNLTIAVSG	626
	: :	
Db	567 LMSSTFTNSYCTTDNLDECLGNQILEVYG	595

RESULT 8

T08934

hypothetical protein F27G19.20 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Mar-2000

C;Accession: T08934

R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.

submitted to the Protein Sequence Database, May 1999

A;Reference number: Z16519

A;Accession: T08934

A;Molecule type: DNA

A;Residues: 1-635 <BEV>

A;Cross-references: EMBL:AL078467; GSPDB:GN00062; ATSP:F27G19.20

A;Experimental source: cultivar Columbia; BAC clone F27G19

C;Genetics:

A;Gene: ATSP:F27G19.20

A;Map position: 4

A;Introns: 38/3; 253/1; 304/1; 414/3

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 18.6%; Score 651.5; DB 2; Length 635;

Best Local Similarity 31.1%; Pred. No. 1.4e-43;

Matches 191; Conservative 104; Mismatches 240; Indels 79; Gaps 19;

```
Qy      30 SESDNSLYFTY---SGQPNTLEVRDLNYQVDLASQVPWFELAQFKMPWTSPSCQNSCEL 86
      : | || |: : | ||: :| | | | : :
Db      17 TNDDRSLPFSIFKKANNPVTLKFENLVYTVKLKDSQGCF-----GKNDKTEERT 65

Qy      87 GIQNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGGKIK-SGQIWINQPPSPQLVR 145
      :: |: |:|::||:| || |: || : || | | | : | | : |:
Db      66 ILKGLTGIVKPGEILAMLGPSGSGKTSLLTALGGRVGEKGKLTGNISYNNKPLS-KAVK 124

Qy     146 KCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRV 205
      : | | : | |||| || | | :|| :| : :: |: : |: | | | : || :
Db     125 RTTGFTVTDALYPNLTVTETLVFTALLRLPNSFKKQEKIKQAKAVMTELGLDRCKDTII 184

Qy     206 GNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTTAHNLVKTLSRLAKGNRLV 265
      | :||:|||||:||||| :| | | :| | ||||| || | : | | | :| | |
Db     185 GGFPLRGVSGGERKRVSIGQEILINPSLLFLDEPTSGLDSTTAQRIVSILWELARGGRTV 244

Qy     266 LISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGY-PCPRYSNPADFVVDL 324
      : ::||| : | | : | : : || :|| | ||:| | :| :
Db     245 VTTIHQP-----SKGNPVYFGLGSNAMDYFASVGYSPLVERINPSDFLLDI 290

Qy     325 TSIDR-----RSREQELATREKAQSLAALFLEKVRDLDDFLWKAETKDLEDTC---V 374
      : | : : | : | : | | : : : : | | |
Db     291 ANGKPLLVISCWPSVGSDESQRPEAMKAALVAFYKTNLLDSVINEVKGD---DLCNKPR 347

Qy     375 ESSVTPLDTNCLPS-PTKMPGAVQQFTTLIRRQIS----NDFRDLPTLLIHGAELMSM 429
      ||| : || || || |::| : : | : | : :|
Db     348 ESS--RVATNTYGDWPTTW---WQQFCVLLKRGKQRRHDSFSGMKV-----AQIFIVSF 397

Qy     430 TIGFLYFGHGSIQLS-FMDTAALLFMIGALIPFNVILDVISKYSERAMLYYELEDGLYT 488
      | |:: :| | || | : | : | |||| | | :|
Db     398 LCGLLWW---QTKISRLQDQIGLLFFISSFWAFFPLFQQIFTFPQERAMLQKERSSGMYR 454

Qy     489 TGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVLVVFCCRIMALA 548
      ||| ::::|:| | : : |||:| | | : |: : | : ||
Db     455 LSPYFLSRVVGDLPMELILPTCFLVITYWMAGLNHNLANFFVTLLVLLVHVLVSGGLGLA 514

Qy     549 AAALLPTFHMASFNSALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCFEGLMKIQF 608
      ||: | : : :| |||: : | | :|| : : :
Db     515 LGALVMDQKSATTLGSMVIMLTFLLAGGYVQ-----HVPVFISWIKY-----VSI 559
```


Qy 609 SRRTYK-MPLGNLT 621
||| : || |
Db 560 GYYTYKLLILGQYT 573

RESULT 9

G02068

white homolog - human

C;Species: Homo sapiens (man)

C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 02-Feb-2001

C;Accession: G02068

R;Croom, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.; Arciniegas, S.; Son, D.; Wu, R.

submitted to the EMBL Data Library, August 1995

A;Reference number: H00769

A;Accession: G02068

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-638 <CRO>

A;Cross-references: EMBL:U34919; NID:g1314276; PIDN:AAC51098.1; PID:g1314277

C;Genetics:

A;Gene: white

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

C;Keywords: ATP; nucleotide binding; P-loop

F;61-253/Domain: ATP-binding cassette homology <ABC>

F;78-85/Region: nucleotide-binding motif A (P-loop)

Query Match 17.6%; Score 618; DB 2; Length 638;
Best Local Similarity 25.7%; Pred. No. 6.5e-41;
Matches 173; Conservative 130; Mismatches 266; Indels 104; Gaps 18;

Qy 33 DNSLYFT--YSGQPN----TLEVRDLNYQVDLASQVPWFELAQFKMPWTSPSCQNSCEL 86
||:| :| | :| ||:| | : ||: : :
Db 17 DNNLTEAQRFFSLPRRAAVNIEFRDLSYSV---PEGPWWRRKKGYKTL----- 60

Qy 87 GIGNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPSQPQLVRK 146
:: :| | ||::||:| || ||::||::: | | | : ||| | : ||
Db 61 -LKGISGKFNSGELVAIMGPSGAGKSTLMNLAGYRETG--MKGAVLINGLRDLRCFRK 117

Qy 147 CVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVG 206
:: | : |||:||||: | : ||:| | : ||::: | | ||:|| |
Db 118 VSCYIMQDDMLLPHLTVQEAMMVSAHLKLQE--KDEGRREMVKEILTALGLLSCANTRTG 175

Qy 207 NMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLRLAKGNRLVL 266
: ||||:|:|:| :||: || : ||||| | : :| : ||:| | : :
Db 176 S-----LSGGQRKRLAIALELVNNPPVMFFDEPTSGLDASCFQVVSMLMKGLAQGGRSII 230

Qy 267 ISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTS 326
::||| : :| ||| : :|| :| | :|| | :| || | ||||| :||: |
Db 231 CTIHQPSAKLFELFDQLYVLSQGQCVRGKVCNLPYLRDLGLNCPYHNPADFVMEVAS 290

Qy 327 IDRRSREQEL--ATREKAQSLAALFLEKVRDL-----DDFLWKAET-----KDLD 369
: : | | || : ||| : ||| : : | |
Db 291 GEYGDQNSRLVRVRE----GMCDSDHKRDLGGDAEVNPFLLWHRPSEEVKQTKRLKGLR 345

Qy 370 EDTCVESVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGAEACIMSM 429
:|: :|| || | :| : || : : : :

Db 346 KDSSSMEGCHSFSASCL-----TQFCILEFKRTFLSIMRDSVLTHLRITSHIGIGL 395
 Qy 430 TIGFLYFGHGSIQLSFMDTAALLF-----MIGALIPFNVILDVISKCYSERAMLYYELE 483
 || || | |: : : || | ||:| :| : ||
 Db 396 LIGLLYLIGIGNEAKKVLNSGFLFFSMLFLMFAALMP-----TVLTFPLE 440
 Qy 484 DGL-----YTTGPHYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFL 534
 |: |: |: || : ::| : : | ||: : |:| |
 Db 441 MGVFLREHLNYWYSLKAYYLAKTMADVFPQIMFPVAYCSIVYWMTSQPSDAVAFVLFAAL 500
 Qy 535 VWLVVFCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVS 594
 : : : | | : : |:| | || : : : | |:| :|
 Db 501 GTMTSLVAQSLGLLIGAASTSLQVATFVGPTAIPVLLFSGFFVSFDTIPTYLQWMSYIS 560
 Qy 595 FLRWCFEGLMKIQF--SRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSG 652
 ::|: |||: : | : : || : :: : || : ||:|
 Db 561 YVRYGFEGVILSIYGLDREDLHCDIDETCHFQKSEAILRELDVENAKLY-LDFIVLG--- 616
 Qy 653 GFMVLYYVSLRFI 665
 : ::||| |
 Db 617 ----IFFISLRLI 625

RESULT 10

D96553

hypothetical protein F5D21.6 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C;Accession: D96553

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;
 Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;
 Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;
 Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;
 Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;
 Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;
 Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-
 Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,
 S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,
 M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;
 Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,
 S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;
 Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: D96553

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-687 <STO>

A;Cross-references: GB:AE005173; NID:g10092349; PIDN:AAG12758.1; GSPDB:GN00141

C;Genetics:

A;Gene: F5D21.6

A;Map position: 1

C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein
F12L6.1; ATP-binding cassette homology

```
Query Match          17.0%; Score 595; DB 2; Length 687;
Best Local Similarity 26.9%; Pred. No. 4.8e-39;
Matches 174; Conservative 127; Mismatches 247; Indels 98; Gaps 19;

Qy      88 IQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINGQPSSPQLVRKC 147
      : | : | : : | : | : : | : | : : | : : | : : | : : |
Db      45 LDGLNGHAEPGRIMAIMGPSGSGKSTLLDSLAGRLARNVIMTGNLLLNKKKA--RLDYGL 102

Qy     148 VAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGN 207
      ||:| | : | : |||||: : | :|| : : : | | | | : ||| :||
Db     103 VAYVTQEDILMGTTLTVRETITYSAHLRLSSDLTKEEVNDIVEGTIIELGLQDCADRIGN 162

Qy     208 MYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHNLVKTL SRLAK-GNRLVL 266
      : ||:|||||:||||: ::| | | | | | | | | :| ::: | :|: | | :|
Db     163 WHSRGVSGGERKRVSVALEILTRPQILFLDEPTSGLDSASAFFVIQALRNIA RDGGRTVV 222

Qy     267 ISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTS 326
      |:||| |:| | | : |:|| :| | :: |:| |:||| |:| : :
Db     223 SSIHQPSSEVFALFDDLFLSSGETVYFGESKFAVEFFAEAGFPCPKRNP SDHFLRCIN 282

Qy     327 ID-----RRSREQELATREKAQSLAALFLEKVRDLDDF----LWKAETKDLDDED 371
      | : | | | : : | : | : | : | : | : | : | : |
Db     283 SDFDVTATLKGSQRI RETP-ATSDPLMNLATSEI-KARLVENYRRSVYAKSAKSRI REL 340

Qy     372 TCVE-----SSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDL----PTL 417
      :| : | : | : | : | : | : | : | : | : | : |
Db     341 ASIEGHHGMEVRKGSEAT-----WFKQLRTLTKRSFVNMC RDIGYYWSRI 385

Qy     418 LIHGAEACILMSMTIGFLYF--GHGSIQLSFMDTAALLFMIGALIPFNVILDV--ISKCYS 473
      :|: : :| :| : : || :| : | : | : | : :
Db     386 VIY----IVVSFCVGTIFYDVGH----SYTSILARVSCGGFITGFMTFMSIGGFPSFIE 436

Qy     474 ERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFL 533
      | : | | | | : | : | : | : | : | : | : | : | : |
Db     437 EMKVIFYKERLSGYYGVSVYIISNYVSSFPFLVAIALITGSITYNMVKFRPGVSHWAFFCL 496

Qy     534 LVWL VVFCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSL----WTVPAW 589
      :: | : : |:|:| | | : : : | | | | | | | |
Db     497 NIFFSVSVIESIMMVVASLVNPNFLMGLITGAGIIGIIMMTSGFFRLLPDL PKVFWRYP-- 554

Qy     590 ISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTI AVSGDKILSAM-----ELDSYPLYAI 643
      || :|: | :| | | : | : :|:|:|: : : : | | |
Db     555 ISFMSYGSWAIQ GAYKNDFLGLFED-PMFAGEPKMTGEQVINKIFGVQVTHSKWWDLSAI 613

Qy     644 YLIVIGLSGGFMVLYYVSLRF-----IKQKPS 670
      ||:: : :|:|: | : :|:|:|
Db     614 VLILV----CYRILFFIVLKLKERAEPALKAIQAKRTMKSLKKRPS 655
```

RESULT 11

T47652

ABC transporter-like protein - Arabidopsis thaliana

N;Alternate names: protein T26I12.10

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
 C;Accession: T47652
 R;Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.;
 Mayer, K.F.X.; Quetier, F.; Salanoubat, M.
 submitted to the Protein Sequence Database, February 2000
 A;Reference number: Z24471
 A;Accession: T47652
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-725 <MON>
 A;Cross-references: EMBL:AL132954
 A;Experimental source: cultivar Columbia; BAC clone T26I12
 C;Genetics:
 A;Map position: 3
 A;Note: T26I12.10
 C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein
 F12L6.1; ATP-binding cassette homology

Query Match 16.9%; Score 591; DB 2; Length 725;
 Best Local Similarity 25.5%; Pred. No. 1.1e-38;
 Matches 175; Conservative 135; Mismatches 273; Indels 104; Gaps 18;

Qy	44	PNTLEVRDLNYQVDLASQVPWF	EQLAQFKMPWTS	SPSCQNSCELGIQNL	SFKVRSGQMLAI	103
		: :	:	:	: : : : :	
Db	70	PYVLNFN	NLQYDVT	LRRRFGF	-----SRQNGVK	TLDDVS
						GEASDGDILAV 115
Qy	104	IGSSGCGRASLLD	VITGRGHGGKIKSGQ	IWINGQP	-SSPQLVRKCV	VAHVRQHNQLLPNLT 162
		: : :	:	: : :	: : :	
Db	116	LGASGAGKSTL	DALAGRVAEGSLR	-GSVTLNGEKVL	QSRLKVISAYVM	QDDLLFPMLT 174
Qy	163	VRETLAFIAQMRL	PRTFSQAQRDKRVED	VIAELRLRQCAD	TRVGNMYVRGL	SGGERRRVS 222
		: :	: : : :	: : :	: : : :	
Db	175	VKETLMFASEFRL	PRSLSKSKMERVEAL	IDQLGLRNAANT	VIGDEGHRGV	SGGERRRVS 234
Qy	223	IGVQLLWNP	GILILDEPTSGLD	SFTAHLNVLK	TL SRLAKGNRL	VLISLHQPRSDIFRLFDL 282
		: : :		: : :	: : : : :	
Db	235	IGIDIHDP	IVLFLDEPTSGLD	STNAFMVVQVL	KRIAQSGSIVIM	SIHQPSARIVELLDR 294
Qy	283	VLLMTSGTPIYL	GAAQHMVQYFTA	IGYPCPRYSNP	ADFYVDLTSID	RRSREQELATREKA 342
		: : : : :	: : : :	:	: : :	
Db	295	LIILSRGKSV	FNGSPASLP	GGFFSDFGRPI	EKENISEFALDLV	-----RELE-GSNEG 347
Qy	343	QSLAALFLEKVR	DLDDFLWK	-----AETKD	LDED-----TC	VESSVTP---LDT 383
		: : :	: : :	: : :	: : : :	
Db	348	KALVD-FNEK	-----WQQNK	ISLIQSAPQ	TNKLDQDRSL	SLKEAINASVSRGKLVS 398
Qy	384	NCLPSPTKM	-----PGAVQ	QFTTLIR	QISNDFRDLPT	LLIHGAELMSMTIGFLY 435
		: : :	: : :	: : :	: : :	
Db	399	SSRSNPTSMET	VSSYANPSLF	FETF-ILAKRYM	KNWIR-MPELV	--GTRIATVMVTGCLLA 454
Qy	436	FGHGSIQLS	FMDTAALLFMIGALIP	--FNVILDV	ISKYSERAMLY	ELEDGLYTTGPYF 493
		: : : : :	: : : :	: : :	: : :	
Db	455	TVYWKLDHT	PRGAQERLTLFA	FVVPTMFYCCL	DNVPVFIQERY	IFLRETHNAYRTSSYV 514
Qy	494	FAKILGELPEH	CAYIIIIYGMP	TYWLANLRPGL	QPFLLHFLLV	VVFCRIMALAAAALL 553
		: : :	: : : :	: : :	: : : :	
Db	515	ISHSLVSLP	QLLAPSLVFS	AITFWTVGL	SGGLEGFV	FYCLLIYASFWSGSSVVTFISGVV 574

```

Qy      554 PTFHMAFFFSNALYNSFYLAGGFMINLSSL---WTVPAWISKVSFLRWCFEGLMKIQFS- 609
      | : | | : | | : | :: : | :: : |
Db      575 PNIMLCYMVSIITYLAYCLLLSGFYVNRDRIPFYWT---WFHYISILKYPYEAVLINEFDD 631

Qy      610 -----RRTYKMPLGNLTIAVSGDKILSAMELDS 637
      : : : | : | : :
Db      632 PSRCFVRGVQVFDSTLLGGVSDSGKVKLETLTSKSLR TKITESTCLRTGSDLLAQQGITO 691

Qy      638 YPLYAIYLVIGLSGGFMVLYYVSLRF 664
      : | | : | : | |
Db      692 LSKWDCLWITFASGLFFRILFYFALLF 718

```

Query Match 16.8%; Score 590; DB 2; Length 559;
Best Local Similarity 29.3%; Pred. No. 9e-39;
Matches 159; Conservative 98; Mismatches 231; Indels 54; Gaps 11;

Qy 266 LISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLT 325
 :|::||| | :: ||: | || | |||| | | |||| | |||| :
 Db 190 IITIHQPSSHVYSLFNNVCLMACGRVIYLGPGDQAVPLFEKCGYPCPAYNPADHLIRTL 249

Qy 326 SIDRRSREQELATREKAQSLAALFLEKVRDLDDFLWKAETKDLEDTC-----VES 376
 :: | : | : |:: | | : || : || : |
 Db 250 AVIDSDRATSMKT-----ISKIR--QGFL----STDLGQSVLAIGNANKLRAAS 292

Qy 377 SVTPLDTNCLPSPTKM-----PGAVQQFTTLIRRQISNDFRDLPTLLIHGAEACIMSM 429
 || ||: || || | | || | : : :
 Db 293 FVTGSDTS---EKTKTFFNQDYNASFWTQFLALFWRSWLTVIRDPNLLSVRLQLILITAF 349

Qy 430 TIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDV-----ISKCYSERAMLYYE 481
 |::| | | : | :: || | :: : :| :: |
 Db 350 ITGIVFF-----QTPVTPATIISINGIM-FNHIRNMNFMLQFPNVPVITAELPIVLRE 401

Qy 482 LEDGLYTTGPGYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVWLWVFC 541
 :|:| | || || : ||:: |:: | | : | :
 Db 402 NANGVYRTSAYFLAKNIAELPQYIILPILYNTIVYWMSGLYPNFWNYCFASLVITILITNV 461

Qy 542 CRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCFC 601
 :: | | : :| || | || | : : :| :| : :|
 Db 462 AISISYAVATIFANTDVAMTILPIFVVPIMAFGGFFITFDAIPSYFKWLSSLSYFKYGYE 521

Qy 602 GL 603
 |
 Db 522 AL 523

RESULT 13

T47648

ABC transporter-like protein - Arabidopsis thaliana

N;Alternate names: protein T15C9.80

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000

C;Accession: T47648

R;Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24470

A;Accession: T47648

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-720 <MEW>

A;Cross-references: EMBL:AL132970

A;Experimental source: cultivar Columbia; BAC clone T15C9

C;Genetics:

A;Map position: 3

A;Note: T15C9.80

C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein

F12L6.1; ATP-binding cassette homology

Query Match 16.8%; Score 589.5; DB 2; Length 720;

Best Local Similarity 25.0%; Pred. No. 1.4e-38;

Matches 173; Conservative 122; Mismatches 289; Indels 109; Gaps 14;

Qy 44 PNTLEVRDLNYQVDLASQVPWFELQAQFKMPWTSPSCQNSCELGIQNLSFKVRSQGMLAI 103
 | | :| | | : :: : : :|| | : | : |::| :|::|:

Db 54 PFVLSFNNLTYNVSVRRKLD FHD-----LVPWRRTSFSKTKTL-LDNISGETRDGEILAV 107

Qy 104 IGSSGCGRASLLDVITGRGHGGKIKSGQIWINGQPSSPQLVRKCVAHVRQHNQLLPNLTV 163
 :|:| |:::|:| : | | :| | :||: :::: |:| | :| | |||

Db 108 LGASGSGKSTLIDALANRIAKGSLK-GTVTLNGEALQSRMLKVISAYVMQDDLLFPMLTV 166

Qy 164 RETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSI 223
 ||| | |: ||||: :::: ||: |: |:| :| | | :|: ||:|||||||

Db 167 EETLMFAAEFRLPRSLPKSKKKLRVQALIDQLGIRNAAKTIIGDEGHRGISGGERRRVSI 226

Qy 224 GVQLLWNPGLILDEPTSGLDSTAHNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLV 283
 |: :: :| :| ||||| ||| :| :|| | |:|: ::::|:| | :| | :

Db 227 GIDIHDPIVLFLDEPTSGLDSTSAFMVVKVLKRIAESGSIIMSIIHQPSHRVLSLLDRL 286

Qy 284 LLMTSGTPIIYLGAQHMVQYFTAIGYPCPRYSNPADFYVDLTS-----IDRRS 331
 : :: | :: |: : :| | | | :| :|| : :

Db 287 IFLSRGHTVFSGSPASLPSFFAGFNPENENQTEFALDLIRELEGSAGGTRGLVEFNK 346

Qy 332 REQELATREKAQSLAA-----LFLEKVRDLDDFLWKAETKDLDDETCVE-----SSVT 379
 : ||: : |:| | | :| : : | |||

Db 347 KWQEMKKQSNPQTLTPPASPNPNLTLK-----EASISIRGKLVSGGGGSSVI 396

Qy 380 PLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGAEACLSMTIGFLY---- 435
 | | : || || | | | || | : :| ||:

Db 397 NHGGGT LAVPAFANPFWIEIKTLTRRSILNSRRQ-PELL--GMRLATVIVT-GFILATVF 452

Qy 436 -----FGHGSIQLSFMDTAAALFMIGALIPFNVILDVISKCYSERAMLYYEL 482
 | : :| | | | : || : |

Db 453 WRLDNSPKGVQERLGFFAFAMSTM-----FYTCADALPVFLQERYIFMRET 498

Qy 483 EDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFL LHFLVWL VVFCC 542
 | | : : | : : :|:| | | || : |: : :

Db 499 AYNAYRRSSYVLSHAIVTFPSLIFLSLAFAVTTFWAVGLEGGLMGFLFYCLIILASFWSG 558

Qy 543 RIMALAAAALLPTFHMAFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFEG 602
 : ::| : |: | | || || : | :| :::|

Db 559 SSFVTFLSGVVPHVMLGYTIVVAILAYFLLFSGFFINRDRIPQYWIWFHYLSLVKYPYEA 618

Qy 603 LMKIQFSRRTY-----KMPLGNLTIAV-----SGDKI 629
 ::: :|| | ||| || : :| :

Db 619 VLQNEFSDPTECFVRGVQLFDNSPLGELTYGMKLRLLDSVSRSIGMRISSTCLTTGADV 678

Qy 630 LSAMELDSYPLYAIYLVIGLSGGFMVLYYVSL 662
 | : : || :| | :|:| :

Db 679 LKQQGVTLQSKWNCLLITVGFGLFRILFYLC 711

RESULT 14

T47650

ABC transporter-like protein - Arabidopsis thaliana

N;Alternate names: protein T15C9.110

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000

C;Accession: T47650

R;Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24470

A;Accession: T47650
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-708 <MEW>
 A;Cross-references: EMBL:AL132970
 A;Experimental source: cultivar Columbia; BAC clone T15C9
 C;Genetics:
 A;Map position: 3
 A;Note: T15C9.110
 C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein
 F12L6.1; ATP-binding cassette homology

Query Match 16.7%; Score 586.5; DB 2; Length 708;
 Best Local Similarity 26.8%; Pred. No. 2.4e-38;
 Matches 187; Conservative 124; Mismatches 254; Indels 133; Gaps 20;

```

Qy      44 PNTLEVRDLNYQVDLASQVPWFEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAI 103
      | | :|:| | | : | | : | : : : | | :|:|:
Db      60 PFLLSFNNLSSYNVLRRRF-----DFSRRKTA-----SVKTLDDITGEARDGEILAV 107

Qy     104 IGSSGCGRASLLDVITGRGHGGKIKSGQIWINGQP-SSPQLVRKCVAVHRQHNQLLPNLT 162
      :| || |::|:| : || :| | :||: :|:: |::| | | ||
Db     108 LGGSGAGKSTLIDALAGRVAEDSLK-GTVTLNGEKVLQSRLLKVISAYVMQDDLLFPMLT 166

Qy     163 VRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVS 222
      |::|| | : : |||: :::: :||| :| :| || ||| :|: ||:|||||||
Db     167 VKETLMFASEFRLPRSLPKSKMERVETLIDQLGLRNAADTVIGDEGHRGVSGGERRRVS 226

Qy     223 IGVQLLWNPGILILDEPTSGLDSTAHNLVKTLSRLAKGNRLVLISLHQPRSDIFRLFDL 282
      ||: :: :| :| ||||| ||| | :|: | |:: :|::|:| || :| | |
Db     227 IGIDIHDPILLFLDEPTSGLDSTNAFMVVQVLKRIAQSGSVVMSI HQPSARIIGLLDR 286

Qy     283 VLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKA 342
      :::: | : : | : :|:: | | | | :| :|: || | :
Db     287 LIILSHGKSVFNGSPVSLPSFFSSFGRIPEKENITEFALDVI-----RELEGSS---- 336

Qy     343 QSLAALFLEKVRDLDDF--LW-----KAETK---DLDEDTCV-----ESSVTPL 381
      | ||| :| | :| | :| | | | | | | | | | | |
Db     337 -----EGTRDLVEFNEKWQQNQATARATTQSRVSLKEAIAASVSRGKLVSGSSGANPI 388

Qy     382 DTNCLPSPTKMPGAVQQFTTLIRRQISNDFRD-----LPTLLIHGAEACILMSMTI--- 431
      : | | : : | :| | | | : |::: | | : | :
Db     389 SMETVSSYANPP--LAETFILAKRYIKNWIRTPELIGMRIGTVMVTG----LLLATVYWR 442

Qy     432 -----GFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLY 479
      || || : : | | | | | | | | | | | |
Db     443 LDNTPRGAQERMGFFAFGMSTM-----FYCCADNIPVFIQERYIFL 483

Qy     480 YELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPYWLANLRPGLQPFLHFLLVWLVV 539
      | | | | : | ||: | | : | : | | | : | :|:
Db     484 RETTHNAYRTSSYVISHALVSLPQLLALSIAFAATTFTWTVGLSGGLESFFYCLIYAAF 543

Qy     540 FCCRIMALAAAALLPTFHMAFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWC 599
      : : : |:| | : : | : | || | | : | :| :|:
Db     544 WSGSSIVTFISGLIPNVMSYMVTIAYLSYCLLLGGFYINRDRIPLYWIWFHYISLLKYP 603

Qy     600 FEGLMKIQF---SR---RTYKMPLGNLTIAVS-----GDKILSAMELDSY 638
  
```



```

      :| :: :| || : :: || || | || : | :
Db      604 YEAVLINEFDDPSRCFVKGVQVFDGTLLEAVSHVMKVKLLDTLSGSLGTKITESTCLRTG 663
Qy      639 P-----LYAIYLIVIGLSGG--FMVLYYVSLRF 664
      |           | : | | : | | :| :| | |
Db      664 PDLLMQQGITQLSKWDCLWITLAWGLFFRILFYLSLLF 701

```

RESULT 15

T31958

hypothetical protein F02E11.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Jan-2000

C;Accession: T31958

R;Favello, A.; Scheet, P.

submitted to the EMBL Data Library, July 1997

A;Description: The sequence of *C. elegans* cosmid F02E11.

A;Reference number: Z21104

A;Accession: T31958

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-658 <FAV>

A;Cross-references: EMBL:AF016661; PIDN:AAB66050.1; GSPDB:GN00020; CESP:F02E11.1

A;Experimental source: strain Bristol N2; clone F02E11

C;Genetics:

A;Gene: CESP:F02E11.1

A;Map position: 2

A;Introns: 115/3; 158/3; 214/3; 330/3; 368/2; 448/3; 525/1

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

```

Query Match          16.5%; Score 579; DB 2; Length 658;
Best Local Similarity 28.3%; Pred. No. 8.5e-38;
Matches 180; Conservative 113; Mismatches 257; Indels 86; Gaps 22;

```

```

Qy      78 PSCQNSCELGIQNLSFKV----RSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIW 133
      | | | | : | | | :| :| :| | | :| :| :| :|
Db      62 PECLAVCALPTSSYQISVSGVAEPGEVLALMGSGAGKTTLMNILAHLDTNGVEYLGVDVT 121

Qy      134 INGQPSSPQLVRKCVAVHRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIA 193
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      122 VNGKKITKQKMRQMCAYVQVDFLCGLTVREQLTYTAHMRMKNATVQ-QKMERVERNVLR 180

Qy      194 ELRLRQCADTRVG-NMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLNV 252
      :: | | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      181 DMNLTDCQNTLIGIPNRMKGISIGEKRLAFACEILTDPKILFCDEPTSGLDAFMASEVV 240

Qy      253 KTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPPIYLGAQHMVQYFTAIG--YP 310
      : | || : :: :| || | :| :| | | :| :| :| :| :| :|
Db      241 RALLDLANKGKTIIVVLHQPSSTVFRMFHKVCFMATGKTVYHGAVDRLCPFFDKLGPDFR 300

Qy      311 CPRYSNPADFYVDLTSIDRRSREQELATR-----EKAQSLAALFLEKVRD-LDDFLWK 362
      | | || | : || : || | : | : | : | : | : | : |
Db      301 VPESYNPADFVMSEISISPET-EQEDVTRIEYLIHEYQNSDIGTQMLKKTRTAVDEFGGY 359

Qy      363 AETKDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGA 422
      : :| | : | | | :| :| :| :| :| :| :| :|
Db      360 GDEDDGESRYNSTFGT-----QFEILLKRSRLRTTFRDPLLLRVREA 401

```

Qy	423	EACLSMTIGFLYFGHGSIQ---	SFMDTAALLFMIGALIPFNVI	LDVISKYSERAMLY	479
	:	:	:	:	:
Db	402	QILATAILVGIV---	NWRVELKGPTIQNLEGVMYNCARDMTFLFYFPSVNVITSELPVFL		458
Qy	480	YELEDGLYTTGPYFFAKIILGELPEHCAYIIIIYGMP	TYWLANLRPGLQPFLLH---	FLLV	535
	:	:	:	:	:
Db	459	REHKSNIYSVEAYFLAKSLAELPQYTILPMIYGTII	YWMAGLVASVTSFLVVFVVCITLT		518
Qy	536	WLVVCCRIMA--LAAAALLPTFHMASF	SNALYNSFYLAGGFMINLSSLWTVPAWISKV		593
	:	:	:	:	:
Db	519	WVAVSIAYVGACIFGDEGLVVTM--	MPMFVLPML-----VFGGFVYNANS---	IPVYYQYV	569
Qy	594	SFLRWC---FEGLMKIQFSRRTYKM-----	PLGNLTI-----AVSGDKILSAMELDSYP		639
	:	:	:	:	:
Db	570	SFVSWFKHGFEALEANQW-KEIDKISGCDLINPLN	ATTGYCPASDGPILTRRGIDT-P		627
Qy	640	LYAIYLI-----VIGLSGGFMVLYYVSLRFIK			666
	:	:	:	:	:
Db	628	LYANVLILFMSFFVYRIIGL-----	VALKIRVRFAK		658

Search completed: February 27, 2004, 07:18:57
Job time : 16.9951 secs

OM protein - protein search, using sw model

Run on: February 27, 2004, 07:17:39 ; Search time 30.2443 Seconds
(without alignments)
4698.604 Million cell updates/sec

Title: US-09-989-981A-8
Perfect score: 3506
Sequence: 1 MAGKAAEERGLPKGATPQDT.....FMVLYYVSLRFIKQKPSQDW 673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query					
No.	Score	Match	Length	DB	ID	Description

1	3506	100.0	673	10	US-09-989-981A-8	Sequence 8, Appli
2	3502	99.9	673	14	US-10-090-455-7	Sequence 7, Appli
3	2883.5	82.2	672	10	US-09-989-981A-4	Sequence 4, Appli
4	1961	55.9	374	15	US-10-415-378-9	Sequence 9, Appli
5	697	19.9	651	9	US-09-837-992-3	Sequence 3, Appli
6	697	19.9	651	10	US-09-989-981A-6	Sequence 6, Appli
7	697	19.9	651	14	US-10-090-455-6	Sequence 6, Appli
8	688.5	19.6	652	9	US-09-837-992-1	Sequence 1, Appli
9	688.5	19.6	652	10	US-09-989-981A-2	Sequence 2, Appli
10	666	19.0	657	9	US-09-866-866A-14	Sequence 14, Appl
11	656	18.7	663	13	US-10-108-605-245	Sequence 245, App
12	642.5	18.3	655	9	US-09-981-353-35	Sequence 35, Appl
13	642.5	18.3	655	14	US-10-120-687-61	Sequence 61, Appl
14	642.5	18.3	655	15	US-10-405-806-2	Sequence 2, Appli
15	640.5	18.3	655	10	US-09-961-086-1	Sequence 1, Appli
16	640.5	18.3	655	15	US-10-405-806-13	Sequence 13, Appl
17	638.5	18.2	655	9	US-09-866-866A-10	Sequence 10, Appl
18	638.5	18.2	655	14	US-10-090-455-5	Sequence 5, Appli
19	634.5	18.1	655	9	US-09-866-866A-27	Sequence 27, Appl
20	618	17.6	638	13	US-10-072-621-10	Sequence 10, Appl
21	617	17.6	674	14	US-10-090-455-4	Sequence 4, Appli
22	617	17.6	674	16	US-10-429-160-10	Sequence 10, Appl
23	609.5	17.4	545	14	US-10-083-357-1335	Sequence 1335, Ap
24	606	17.3	676	15	US-10-369-493-3799	Sequence 3799, Ap
25	592.5	16.9	1095	15	US-10-369-493-2025	Sequence 2025, Ap
26	590	16.8	559	15	US-10-369-493-5740	Sequence 5740, Ap
27	580	16.5	599	15	US-10-210-130-14	Sequence 14, Appl
28	579	16.5	658	15	US-10-369-493-5347	Sequence 5347, Ap
29	573.5	16.4	627	14	US-10-090-455-8	Sequence 8, Appli
30	572.5	16.3	646	13	US-10-154-452-4	Sequence 4, Appli
31	571.5	16.3	646	14	US-10-090-455-13	Sequence 13, Appl
32	570.5	16.3	646	13	US-10-154-452-8	Sequence 8, Appli
33	567.5	16.2	646	13	US-10-072-621-9	Sequence 9, Appli
34	567.5	16.2	646	14	US-10-090-455-2	Sequence 2, Appli
35	562.5	16.0	1049	15	US-10-369-493-1520	Sequence 1520, Ap
36	558.5	15.9	646	14	US-10-079-087-2	Sequence 2, Appli
37	539	15.4	695	15	US-10-369-493-6199	Sequence 6199, Ap
38	535	15.3	560	15	US-10-369-493-12899	Sequence 12899, A
39	525.5	15.0	551	15	US-10-369-493-3562	Sequence 3562, Ap
40	524	14.9	608	15	US-10-369-493-5748	Sequence 5748, Ap
41	521	14.9	639	15	US-10-369-493-6184	Sequence 6184, Ap
42	518	14.8	604	9	US-09-745-763-197	Sequence 197, App
43	511	14.6	610	15	US-10-369-493-5687	Sequence 5687, Ap
44	475.5	13.6	615	10	US-09-949-029-24	Sequence 24, Appl
45	464	13.2	1564	9	US-09-801-368-244	Sequence 244, App

ALIGNMENTS

RESULT 1
 US-09-989-981A-8
 ; Sequence 8, Application US/09989981A
 ; Publication No. US20030049730A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hobbs, Helen H.

```

; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 673
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human ABCG8 (hABCG8)
US-09-989-981A-8

```

```

Query Match          100.0%; Score 3506; DB 10; Length 673;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MAGKAAEERGLPKGATPQDTSGLQDRLFSSSDNSLYFTYSGQPNTLEVRDLNYQVDLAS 60
|
Db      1 MAGKAAEERGLPKGATPQDTSGLQDRLFSSSDNSLYFTYSGQPNTLEVRDLNYQVDLAS 60
|
Qy     61 QVPWFEEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITG 120
|
Db     61 QVPWFEEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITG 120
|
Qy    121 RGHGGKIKSGQIWINQPPSSPQLVRKCAVHRQHNQLLPNLTVRETAFIAQMRLPRTFS 180
|
Db    121 RGHGGKIKSGQIWINQPPSSPQLVRKCAVHRQHNQLLPNLTVRETAFIAQMRLPRTFS 180
|
Qy    181 QAQRDKRVEDVIAELRLRQCADTRVGNNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPT 240
|
Db    181 QAQRDKRVEDVIAELRLRQCADTRVGNNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPT 240
|
Qy    241 SGLDSFTAHLNVLKTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM 300
|
Db    241 SGLDSFTAHLNVLKTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM 300
|
Qy    301 VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL 360
|
Db    301 VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL 360
|
Qy    361 WKAETKDLDEDTCEVSSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH 420
|
Db    361 WKAETKDLDEDTCEVSSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH 420
|
Qy    421 GAEACLMSMTIGFLYFGHSGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY 480
|

```

```

Db      421 GAEACILMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKYSERAMLYY 480
Qy      481 ELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPYWLANLRPGLQPFLHFLLVWL VVF 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 ELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPYWLANLRPGLQPFLHFLLVWL VVF 540
Qy      541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF 600
Qy      601 EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIIVIGLSGGFMVLYYV 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIIVIGLSGGFMVLYYV 660
Qy      661 SLRFIKQKPSQDW 673
          ||||||||||||
Db      661 SLRFIKQKPSQDW 673

```

RESULT 2

US-10-090-455-7

; Sequence 7, Application US/10090455

; Publication No. US20030027259A1

; GENERAL INFORMATION:

; APPLICANT: Chen, Hongyun

; APPLICANT: Le Bihan, Stephane

; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF

; FILE REFERENCE: 100103.406

; CURRENT APPLICATION NUMBER: US/10/090,455

; CURRENT FILING DATE: 2002-03-01

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 673

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-090-455-7

Query Match 99.9%; Score 3502; DB 14; Length 673;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1 MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLAS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLAS 60
Qy      61 QVPWFEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 QVPWFEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITG 120
Qy      121 RGHGGKIKSGQIWINQGPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 RGHGGKIKSGQIWINQGPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFS 180
Qy      181 QAQRDKRVEDVIAELRLRQCADTRVG NMYVRGLSGGERRRVSIGVQLLWNP GILILDEPT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 QAQRDKRVEDVIAELRLRQCADTRVG NMYVRGLSGGERRRVSIGVQLLWNP GILILDEPT 240

```

Qy	241	SGLDSFTAHNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM	300
Db	241	SGLDSFTAHNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM	300
Qy	301	VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL	360
Db	301	VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL	360
Qy	361	WKAETKDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH	420
Db	361	WKAETKDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH	420
Qy	421	GAEACLSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY	480
Db	421	GAEACLSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY	480
Qy	481	ELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWL VVF	540
Db	481	ELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWL VVF	540
Qy	541	CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF	600
Db	541	CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF	600
Qy	601	EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV	660
Db	601	EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV	660
Qy	661	SLRFIKQKPSQDW	673
Db	661	SLRFIKQKPSQDW	673

RESULT 3

US-09-989-981A-4

; Sequence 4, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

; APPLICANT: Tian, Hui

; APPLICANT: Tularik Inc.

; APPLICANT: Board of Regents, The University of Texas System

; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use

; FILE REFERENCE: 018781-007320US

; CURRENT APPLICATION NUMBER: US/09/989,981A

; CURRENT FILING DATE: 2002-07-23

; PRIOR APPLICATION NUMBER: US 60/252,235

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/253,645

; PRIOR FILING DATE: 2000-11-28

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 672

RESULT 4

US-10-415-378-9

```
; Sequence 9, Application US/10415378
; Publication No. US20040014945A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; TANG, Y. Tom
; APPLICANT: YUE, Henry; NGUYEN, Danniel B.;
; APPLICANT: HAFALIA, April J.A.; ELLIOTT, Vicki S.;
; APPLICANT: LU, Yan; CHAWLA, Narinder K.;
; APPLICANT: YAO, Monique G.; BAUGHN, Mariah R.;
; APPLICANT: GANDHI, Ameena R.; DING, Li;
; APPLICANT: SANJANWALA, Madhusudan M.; RAMKUMAR, Jayalaxmi;
; APPLICANT: ARVIZU, Chandra S.; GIETZEN, Kimberly J.;
; APPLICANT: LAL, Preeti G.; AZIMZAI, Yalda;
; APPLICANT: KHAN, Farrah A.; THANGAVELU, Kavitha;
; APPLICANT: THORNTON, Michael B.; LU, Dyung Aina M.;
; APPLICANT: TRIBOULEY, Catherine M.; WARREN, Bridget A.;
; APPLICANT: ISON, H. Craig; DAS, Debopriya;
; APPLICANT: RAUMANN, Brigitte E.; POLICKY, Jennifer L.;
; APPLICANT: KEARNEY, Liam
; TITLE OF INVENTION: TRANSPORTERS AND ION CHANNELS
; FILE REFERENCE: PI-0270 USN
; CURRENT APPLICATION NUMBER: US/10/415,378
; CURRENT FILING DATE: 2003-05-07
; PRIOR APPLICATION NUMBER: PCT/US01/46055
; PRIOR FILING DATE: 2001-10-27
; PRIOR APPLICATION NUMBER: US 60/250,790
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 60/252,232
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/249,661
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: US 60/247,673
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/245,904
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/243,989
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040014945A1 6585710CD1
US-10-415-378-9
```

```
Query Match          55.9%; Score 1961; DB 15; Length 374;
Best Local Similarity 99.7%; Pred. No. 7.1e-189;
Matches 373; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      300 MVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDF 359
      || |||||
```

Db 1 MVHYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDF 60

Qy 360 LWKAETKDLDEDTCESSVTPLDTNCLPSPKMPGAVQQFTTLIRROIISNDFRDLPTLLI 419
 |||

Db 61 LWKAETKDLDEDTCESSVTPLDTNCLPSPKMPGAVQQFTTLIRROIISNDFRDLPTLLI 120

Qy 420 HGAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLY 479
 |||

Db 121 HGAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLY 180

Qy 480 YELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLLHFLLVWLTV 539
 |||

Db 181 YELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLLHFLLVWLTV 240

Qy 540 FCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWC 599
 |||

Db 241 FCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWC 300

Qy 600 FEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYY 659
 |||

Db 301 FEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYY 360

Qy 660 VSLRFIKQKPSQDW 673
 |||

Db 361 VSLRFIKQKPSQDW 374

RESULT 5

US-09-837-992-3

; Sequence 3, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 651

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)

; OTHER INFORMATION: amino acid sequence

US-09-837-992-3

Query Match

19.9%; Score 697; DB 9; Length 651;

Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

; APPLICANT: Tian, Hui

```
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-6
```

```
Query Match          19.9%; Score 697; DB 10; Length 651;
Best Local Similarity 28.9%; Pred. No. 7.4e-61;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;
```

```
Qy      16 TPQDTSGLQDRLFSSESNDLSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFQQLAQFKMPW 75
      || : ||| | | | : : || : : | | : || : : |
Db      8 TPGGSMGLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQQW 64

Qy      76 TSPSCQNSCELGIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI 134
      | : : : | | ||| : | || | : : || : | | | : : :
Db      65 TRQI-----LKDVSLYVESGQIMCILGSSSGSKT'TLLDAMSGRLGRAGTF-LGEVYV 115

Qy     135 NGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAE 194
      || : : | : | | : || : |||| : | : : | : | : || |
Db     116 NGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAI-RRGNPGSFQKKVEAVMAE 174

Qy     195 LRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSEFTAHNLVKT 254
      | | | | : || : | : | | |||| | || : | : : |||| : || | : |
Db     175 LSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVL 234

Qy     255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314
      | || : || : : : |||| : : || : : : | | : | | : : | |||| :
Db     235 LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEH 294

Qy     315 SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK 362
      ||| ||| : |||| : | : : : | : : | : : : : : : : | : : : |
Db     295 SNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSAICHKTLKNIERMKHL----- 348

Qy     363 AETKDLDEDTCESSVTPLDTNCLPSPTK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHG 421
      | : : : | | | | : | : | | : | : | | : : :
Db     349 -----KTLPM----VPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRL 390

Qy     422 AEACLSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVILDVISKYSER 475
      : : | : : | | : | | | | : | : | : : : |
Db     391 LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DRVGLLYQFVGATPYTGMLNAVNLFPPVLR 446

Qy     476 AMLYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLV 535
```

```

      |: | :||| | | | | :|: | | | : |
Db      447 AVSDQESQDGLYQKWQMLLAYALHVLPLFSVVATMIFSSVCYWTGLGLHPEVARF----- 499
Qy      536 WLVVFFCCRIMALAAAALLPTFHMASFFS-----NALYNSFYLAG-----GFM 577
      :||| : | : |:: :|| | :
Db      500 -----GYFSAALLAPHLIGEFTLVLLGIVQNPINIVNSVVALLSIAGVLVSGGFL 549
Qy      578 INLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVS 625
      |: : | | :| ::| | |: :| : |:: :
Db      550 RNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTN 597

```

RESULT 7

US-10-090-455-6

```

; Sequence 6, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090,455
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-6

```

```

Query Match          19.9%; Score 697; DB 14; Length 651;
Best Local Similarity 28.9%; Pred. No. 7.4e-61;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

```

```

Qy      16 TPQDTSGLDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFEEQLAQFKMPW 75
      || : ||| | | | : :||| : :| | : ||: : : |
Db      8 TPGGSMGLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQQW 64

Qy      76 TSPSCQNSCELGIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI 134
      | :::| | ||:: |:||| |: :||| ::|| | | |:::
Db      65 TRQI-----LKDVS LYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTF-LGEVYV 115

Qy      135 NGQPSSPQLVRKCVAHVRQHNLPLNTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAE 194
      ||: : : | ::| | : || :||| || : | : : | : | :|| |::|
Db      116 NGRALRREQFQDCFSYVLQSDTLLSSLTVRET LHYTALLAI-RRGNPGSFQKKVEAVMAE 174

Qy      195 LRLRQCADTRVGNMYVRGLSGGERRRVSGVQLLWNPGLILDEPTSGLDSTAHNLVKT 254
      | | || :|| : |:| ||||| ||| :| ::| |||: || |:::|
Db      175 LSLSHVADRLIGNYSLGISTGERRRVSIAAQLLQDPKVMLEFDEPTTGLDCMTANQIVVL 234

Qy      255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314
      | ||: ||:::||||:::|||| : ::| | : | : :| |||| :
Db      235 LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFENDCGYPCPEH 294

Qy      315 SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK 362
      ||| |||:||||:| :::| : : : : : :|:::|

```

Db 295 SNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHL----- 348
 Qy 363 AETKDLDEDTCESSVTPLDTNCLPSPTK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHG 421
 Db 349 -----KTLPM---VPFKTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRL 390
 Qy 422 AEACLMSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSER 475
 Db 391 LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DRVGLLYQFVGATPYTGMLNAVNLFPVLR 446
 Qy 476 AMLYYELEDGLYTTGPFYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLLHFLLV 535
 Db 447 AVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARF----- 499
 Qy 536 WLTVFCCRIMALAAAALLPTFHMASFFS-----NALYNSFYLAG-----GFM 577
 Db 500 -----GYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSVALLSIAGVLVSGGFL 549
 Qy 578 INLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVS 625
 Db 550 RNIQEMPIPFKIIISYFTFQKYCSEILVNEFYGLNFTCGSSNVSVTTN 597

RESULT 8

US-09-837-992-1

; Sequence 1, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

; TITLE OF INVENTION: and Methods of Use

; FILE REFERENCE: 018781-006020US

; CURRENT APPLICATION NUMBER: US/09/837,992

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: US 60/198,465

; PRIOR FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/204,234

; PRIOR FILING DATE: 2000-05-15

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 652

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)

; OTHER INFORMATION: amino acid sequence

US-09-837-992-1

Query Match 19.6%; Score 688.5; DB 9; Length 652;

Best Local Similarity 28.1%; Pred. No. 5.3e-60;

Matches 188; Conservative 125; Mismatches 233; Indels 123; Gaps 16;

Qy 45 NTLEVRDLNYQVDLASQV-PWFEQLAQFKMPWTSPSCQNSCELGI-QNLSFKVRSGQMLA 102

```

      ::| | ::| | :::| | | | | | | : | ::| : ||::
Db      37 HSLGVLHVSYSV--SNRVGPW-----WNIKSCQQKWDRQILKDVSLYIESGQIMC 84

Qy      103 IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINGQPSSPQLVRKCVAHVRQHNQLLPNLT 162
      |:|||| |: :|| |:|| |:::| | : | :| | :| |
Db      85 ILGSSGSGKTTLLDAISGRLLRRTGTLEGEVFNVCCELRRDQFQDCFSYVLQSDVFLSSLT 144

Qy      163 VRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVS 222
      ||||| : | : | | : | :|:| | : || | | :|: | :| |||||
Db      145 VRETLRYTAMLALCRS-SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVS 203

Qy      223 IGVQLLWNPGLILDEPTSGLDSTAHNLVKTLSRLAKGNRLVLISLHQPRSDIFRLFDL 282
      | ||| :| :::|||||:|| | | :| | :| :|:|:|:|:|:|:| |
Db      204 IAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFQDK 263

Qy      283 VLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKA 342
      : ::| | : : | : | : | | | | : || | | :|:|:|:| :|:|:| : :
Db      264 IAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRV 323

Qy      343 QSLAALF-----LEKVRDLDDFLWKAETKDLEDTCESSVTPLDTCNCLPSPT 390
      | | | | : | : | | | | | : | : | |
Db      324 QMLECAFKESDIYHKILENIERARYL-----KTLPM----VPFKT 359

Qy      391 K-MPGAVQQFTTLIRROI SNDFRDLPTLLIHGAELMSMTIGF--LYFGHSGIQLSFMD 447
      | || : : | | | | : : : : : | : : : : |
Db      360 KDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLKGAVQD 419

Qy      448 TAALLFMIGALIPFNVILDVISKYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAY 507
      ||: : | : :|: : | | : | : || | : | ||
Db      420 RVGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIA 479

Qy      508 IIIYGMPTYWLANLRPGLQPFLLHFLLVLVVFCCRIMALAAAALLPTFHMASFFSNAL- 566
      :|: || | | : | : || | : | : |
Db      480 TVIFSSVCYWTGLGYPEVARF-----GYFSAALLAPHLIGEFLLTVLL 522

Qy      567 -----YNSFYLAGGFMINLSSLWTVPWISKVSFLRWCFEGLMKIQFS 609
      : : ||: | : : : | :| | | : |
Db      523 GIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEF- 581

Qy      610 RRTYKMPGLNLTIAVSGDKILSAMELDSYPLYAI-----YLIVI 648
      | : | | | : | :|:| | : ||:
Db      582 ---YGL---NFTCGGSNTSML-----NHPMCAITQGVQFIEKTCPGATSREFTANFLILY 629

Qy      649 GLSGGFMVL 657
      | : :|
Db      630 GFIPALVIL 638

```

RESULT 9

US-09-989-981A-2

; Sequence 2, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

; APPLICANT: Tian, Hui

```
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-2
```

```
Query Match          19.6%; Score 688.5; DB 10; Length 652;
Best Local Similarity 28.1%; Pred. No. 5.3e-60;
Matches 188; Conservative 125; Mismatches 233; Indels 123; Gaps 16;
```

```
Qy      45 NTLEVRDLNYQVDLASQV-PWFEQLAQFKMPWTSPSCQNSCELGI-QNLSFKVRSGQMLA 102
      ::| | :| | ::| || | || : | ::| : |||::
Db      37 HSLGVLHVSYSV--SNRVGPW-----WNKSCQQKWDRQILKDVSLYIESGQIMC 84

Qy      103 IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPSQPQLVRKCAVHRQHNQLLPNLT 162
      |:|||| |: :||| |:|| |:::| | : | :||
Db      85 ILGSSGSGKTTLLDAISGRRLRTGTLEGEVFVNGCELRRDQFQDCFSYVLQSDVFLSSLT 144

Qy      163 VRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVS 222
      ||||| : | : | | : | :||| | : || | || :| : | |||||
Db      145 VRETLRYTAMLALCRS-SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVS 203

Qy      223 IGVQLLWNPGLILDEPTSGLDSFTAHNVLKTL SRLAKGNRLVLISLHQPRSDIFRLFDL 282
      | ||| :| ::| ||||:|||| ||: :| | : ||: :|:|:|:|:|:|:|:|
Db      204 IAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFDDK 263

Qy      283 VLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKA 342
      : ::| | : : | : | : | | |||| :|| | ||:| |||:| :|||:| : | :
Db      264 IAILTYGELVFCGTPEEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRV 323

Qy      343 QSLAALF-----LEKVRDLDDFLWKAETKDLDEDTCESSVTPLDTNCLPSPT 390
      | | | :| : | | :| : | | : | : |
Db      324 QMLECAFKESDIYHKILENIERARYL-----KTLPM----VPFKT 359

Qy      391 K-MPGAVQQFTTLIRRQISNDFRDLPTLLIHGAEACLSMTIGF--LYFGHGSIQLSFMD 447
      | || : | :| | | : : : : | : : | : : : |
Db      360 KDPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNTLKGAVQD 419

Qy      448 TAALLFMIGALIPFNVILDVISKCYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAY 507
      ||: : | : : : ||: | :||| | : | ||
Db      420 RVGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIA 479

Qy      508 IIIYGMPTYWLANLRPGLQPFLHFLLVVLFCCRIMALAAAALLPTFHMASFFSNAL- 566
```



```

          :|:  ||  | | : |          :|||  : | : |
Db      480 TVIFSSVCYWTGLGLYPEVARF-----GYFSAALLAPHLIGEFLTLVLL 522
Qy      567 -----YNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCFEGLMKIQFS 609
          :  :  ||: | :  :  : | :| | | : :|
Db      523 GIVQNPNIIVNSIVALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEF- 581
Qy      610 RRTYKMPPLGNLTIAVSGDKILSAMELDSYPLYAI-----YLIVI 648
          | :  | |  | :|  :||: ||          :||:
Db      582 ---YGL---NFTCGGSNTSML-----NHPMCAITQGVQFIEKTCPGATSREFTANFLILY 629
Qy      649 GLSGGFMVL 657
          |  :|
Db      630 GFIPALVIL 638

```

RESULT 10

US-09-866-866A-14

; Sequence 14, Application US/09866866A

; Patent No. US20020102244A1

; GENERAL INFORMATION:

; APPLICANT: Sorrentino, Brian

; APPLICANT: Schuetz, John

; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells

; FILE REFERENCE: 1340-1-021CIP2

; CURRENT APPLICATION NUMBER: US/09/866,866A

; CURRENT FILING DATE: 2001-08-30

; PRIOR APPLICATION NUMBER: 09/584,586

; PRIOR FILING DATE: 2000-05-31

; PRIOR APPLICATION NUMBER: PCT/US99/11825

; PRIOR FILING DATE: 1999-05-27

; PRIOR APPLICATION NUMBER: 60/086,988

; PRIOR FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 14

; LENGTH: 657

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-866-866A-14

Query Match 19.0%; Score 666; DB 9; Length 657;

Best Local Similarity 28.0%; Pred. No. 1e-57;

Matches 178; Conservative 127; Mismatches 255; Indels 76; Gaps 14;

```

Qy      91 LSF-----KVRSGQML-----AIIGSSGCGRASLLDVITGRG 122
          |||  ||: || :|          ||: | :| |:: ||||: |
Db      37 LSFHHITYRVKVKSGFLVRKTVEKEILSDINGIMKPGLNAILGPTGGGKSSLLDVLAAR- 95
Qy      123 HGGKIKSGQIWINQGPSSPQLVRKCVAVHRQHNQLLPNLTVRETAFIAQMRLPRTFSQA 182
          |  || : ||| | | : | :| | : :| |||| | | | :||| |
Db      96 KDPKGLSGDVLINGAP-QPAHFKCCSGYVVQDDVVMGTLTVRENLQFSAAALRLPTTMKNH 154
Qy      183 QRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSG 242
          ::::|:  :| || | : ||::||  :||: ||||: | ||:::| :| || ||||: |
Db      155 EKNERINTIIKELGLEKVADSKVGTQFIRGISGGERKRTSIGMELITDPSILFLDEPTTG 214

```

QY 243 LDSFTAHLVKTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQ 302
 ||| ||: :: | |::| | :: |::||| ||: || : | | ::
 Db 215 LDSSTANAVLLLLKRMSKQGRTIIFSIHQPRYSIFKLFDLTLASGKLVFHGPAQKALE 274
 QY 303 YFTAIGYPCPRYSNPADFYVDLTSIDRR----SREQELATREKAQSLAALFLEKVRDLDD 358
 || : || | |::|||::|: : | : : : : : : : :
 Db 275 YFASAGYHCEPYNNPADFFLDVINGDSSAVMLNREEQDNEANKTEEPSKGEKPVLENLSE 334
 QY 359 FLWKA----ETK-DLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRD 413
 | : ||| :||: : : : | : | : | : || | :
 Db 335 FYINSAIYGETKAELDQ---LPGAQEKKGTSFAKPEVYVTSFCHQLRWIARRSFKNLLGN 391
 QY 414 LPTLLIHGAEACILMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISCYS 473
 : : : || :|| : | :|| : : :||:
 Db 392 PQASVAQLIVTVILGLIIGAIYFDLKYDAAGMQNRAGVLFFL-----TTNQCF 440
 QY 474 -----ERAMLYYELEDGLYTTGPYFFAKILGE-LPEHCAYIIYGMPTYWLANLR 522
 |: : :| | | || |:: : || :|: |:: |:
 Db 441 SVSAVELFVVEKKLFIHEYISGYRVSYYFFGKVMSDLPLMRFLPSVIFTCILYFMLGLK 500
 QY 523 PGLQPFLLHFLLVVLFCCRIMALAAAALLPTFHMAFFSNALYNSFYLAGGFMINLSS 582
 : | : : :| : ||| | :|: : : | | :||:
 Db 501 KTVDAFFIMFTLIMVAYTASSMALAIATGQSVSVATLLMTIAFVFMMLFSGLLVNLRT 560
 QY 583 LWTVPWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT-----IAVSGDKIL--S 631
 : :|: | | :| : : | | :|: |:
 Db 561 IGPWLSWLQYFSIPRYGFTALQYNEFLGQEF-CPGFNVTDNSTCVNSYAICTGNEYLINQ 619
 QY 632 AMELDSYPLYAIYLIVIGLSGGFMVLYYVSLRFIKQ 667
 :|| : |: : : : |: : |: | |:::
 Db 620 GIELSPWGLWKNHVALACMIIIFLTIAYLKLLFLKK 655

RESULT 11

US-10-108-605-245

; Sequence 245, Application US/10108605

; Publication No. US20020160934A1

; GENERAL INFORMATION:

; APPLICANT: Broadus, Julie

; APPLICANT: Stam, Lynn

; APPLICANT: Bachmann, Jane

; APPLICANT: Kamdar, Kim

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE

; TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF

; FILE REFERENCE: 31133B

; CURRENT APPLICATION NUMBER: US/10/108,605

; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: US 09/761,142

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/176,418

; PRIOR FILING DATE: 2000-01-14

; NUMBER OF SEQ ID NOS: 361

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 245

; LENGTH: 663

; TYPE: PRT

```
; ORGANISM: Drosophila melanogaster
US-10-108-605-245
```

Query Match		18.7%;	Score 656;	DB 13;	Length 663;
Best Local Similarity		30.3%;	Pred. No. 1e-56;		
Matches		178;	Conservative 113;	Mismatches 265;	Indels 32; Gaps 10;

Qy	88	IQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGG--KIKSGQIWINQQPSSPQLVR	145
		: : : : : : : : : : : : :	
Db	89	LKNVCGVAYPGELLAVMGSSGAGKTTLLNALAFRSPQGIQVSPSGMRLNLNGQPVDAKEMQ	148
Qy	146	KCVAHVRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRV	205
		: : : : : : : : :	
Db	149	ARCAYVQQDDLFIGSLTAREHLIFQAMVRMPRHLTYRQRVARVDQVIQELSLSKCQHTII	208
Qy	206	G-NMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNVLKTL SRLAKGNRL	264
		: : : : : : : : :	
Db	209	GVPGRVKGLSGGERKRLAFASEALTDPPLLICDEPTSGLDSTAHSVVQVLKKLSQKGBT	268
Qy	265	VLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAQHMVQYFTAIGYPCPRYSNPADFYVDL	324
		: : : : : : :	
Db	269	VILTIHQPSSELFELFDKILLMAEGRVAFLTGPSEAVDFFSYVGAQCPTNYPADFYVQV	328
Qy	325	TSIDRRSREQELATREKAQSLAALF-LEKV-RDLDDFLWKAETKDLDEDTCESSVTPLD	382
		: : : : : : : : : : :	
Db	329	LAV---VPGREIESRDRIAKICDNFAISKVARDMEQLL---ATKNLEK-----PLE	373
Qy	383	TNCLPSP---TKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGAEACLSMTIGFLYFGH	438
		: : : : : : : : : :	
Db	374	-----QPENGYTYKATWFMQFRAVLWRSWLSVLKEPLLVKVRLIQTTMVAILIGLIFLGQ	428
Qy	439	GSIQLSFMDTAALLFMIGALIPFNVILDVISKYSERAMLYYELEDGLYTTGPYFFAKIL	498
		: : : : : : : : : :	
Db	429	QLTQVGVMNINGAIFLFLTNTMTFQNVFATINVTSELPVFMREARSRLYRCDTYFLGKTI	488
Qy	499	GELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVVLVVFCCRIMALAAAALLPTFHM	558
		: : : : : : : : : : :	
Db	489	AELPLFLTVPPLVFTAIAYPMIGLRAGVLHFFNCLALVTLVANVSTSFGYLISCASSSTSM	548
Qy	559	ASFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFEGLMKIQFS---RRTYKM	615
		: : : : : : : : :	
Db	549	ALSVGPPVVIIPFLFGGFFLNSGSVPVYLKWLSYLSWFRYANEGLLINQWADVPEGEISC	608
Qy	616	PLGNLTIAVSGDKILSAMELDSYPLYAIYLVIGLSGGFMVLYYVSLR	663
		: : : : : : : : :	
Db	609	TSSNTTCPSSGKVIETLNFSAADLPLDYVGLAILIVSFRVLAYLALR	656

```

RESULT 12
US-09-981-353-35
; Sequence 35, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US

```

; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 35
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1
US-09-981-353-35

Query Match 18.3%; Score 642.5; DB 9; Length 655;
Best Local Similarity 27.2%; Pred. No. 2.4e-55;
Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

```
Qy      19 DTSGQLQDRLFSSSDNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFELAQFKMPWTSP 78
          :|:|      | : : | |      |      | :| :| | |      :|
Db      16 NTNG-----FPATASNDLKAFTEGA--VLSFHNICYRVKLKSGF-----LP----- 54

Qy      79 SCQNSCELGI-QNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQ 137
          |:      | |      |::      :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      55 -CRKPVEKEILSNINGIMKPG-LNAILGPTGGGKSSLLDLVLAARKDPSGL-SGDVLINGA 111

Qy     138 PSSPQLVRKC-VAHVRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELR 196
          |      | |      :| | : :| :| | | | | | :| | : :| :| :| | |
Db     112 PRPANF--KCN SGYVVQDDVVMGTLTVRENLQFS AALRLATMTNHEKNERINRVIQELG 169

Qy     197 LRQCADTRVGNMYVRGLSGGERRRVSGVQLLWNP GILILDEPTSGLDSEFTAHNLVKTLS 256
          | : | | :| | :| :| :| | | | | :| | | | | :| | :| :| :|
Db     170 LDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLK 229

Qy     257 RLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSN 316
          | :| :| | :| :| | | | | | | :| :| :| :| :| :| :| :|
Db     230 RMSKQGRTIIFS I HQPRYSIFKLFD SLTLLASGRLMFHGPAQEALGYFESAGYHCEAYNN 289

Qy     317 PADFYVDLTSIDRR----SREQELATRE--KAQSLAALFLEKVRDL--DDFLWKAETK-- 366
          | | | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     290 PADFFLDIINGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYK-ETKAE 348

Qy     367 -----DLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLI 419
          : :      | : :      :|      | : :|      | : :
Db     349 LHQLSGGEKKKKITVFKEISYTTSF-----HQLRWVSKRSFKNLLGNPQASIA 397

Qy     420 HGAEACILMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKYS----- 473
          : : : | | :| | :      : | :| :      :| :|
Db     398 QIIVTVVLGLVIGAIYFGLKNDSTGIQNRAGVLFLL-----TTNQCFSSVSAVE 446

Qy     474 -----ERAMLYYELEDGLYTTGPYFFAKILGE-LPEHCAYIIIIYGMPTYWLANLRPGLQPF 528
          | : : :| | | | | | | :| :| :| :| :| :| :| :|
Db     447 LFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLPMRMLPSIIFTCIVYFMLGLKPKADAF 506

Qy     529 LLHFLLVVLVVFCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPA 588
          : : :| : | | | | | :| :      : : | :| :| :| :| :|
Db     507 FVMFTLMMVAYSASSMALAIAAGQSVVSVATLLMTICFVFMIFSGLLVNLTITIASWLS 566
```

Qy 589 WISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT-----IAVSGDKIL--SAMELDSYP 639
 |: | |: | | :| : : | | | :|:| | :| :
 Db 567 WLQYFSIPRYGFTALQHNEFLGQNF-CPGLNATGNNPCNYATCTGEEYLVKQGIDLSPWG 625
 Qy 640 LYAIYLIVIGLSGGFMVLYYVSLRFIKQ 667
 |: :: : : |: : |: | |:|:
 Db 626 LWKNHVALACMIVIFLTIAYLKLFLKK 653

RESULT 13

US-10-120-687-61

; Sequence 61, Application US/10120687
 ; Publication No. US20030082155A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Massachusetts General Hospital
 ; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in
 Treating Diabetes
 ; TITLE OF INVENTION: Mellitus
 ; FILE REFERENCE: 3284/1235B
 ; CURRENT APPLICATION NUMBER: US/10/120,687
 ; CURRENT FILING DATE: 2002-04-11
 ; PRIOR APPLICATION NUMBER: US60/169082
 ; PRIOR FILING DATE: 1999-12-06
 ; PRIOR APPLICATION NUMBER: US 09/963,875
 ; PRIOR FILING DATE: 2001-09-25
 ; PRIOR APPLICATION NUMBER: US 60/215109
 ; PRIOR FILING DATE: 2000-06-28
 ; PRIOR APPLICATION NUMBER: US 60/238880
 ; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: US 09/731261
 ; PRIOR FILING DATE: 2000-12-06
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 61
 ; LENGTH: 655
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-120-687-61

Query Match 18.3%; Score 642.5; DB 14; Length 655;
 Best Local Similarity 27.2%; Pred. No. 2.4e-55;
 Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

Qy 19 DTSGQLQDRLFSSSESDNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFQQLAQFKMPWTSP 78
 :|:| | : : | | | : : | | | :|
 Db 16 NTNG-----FPATASNDLKAFTEGA--VLSFHNICYRVKLKSGF-----LP----- 54
 Qy 79 SCQNSCELGI-QNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQ 137
 |: | | |: : : | : ||:| :| |:| |||: | : || : |||
 Db 55 -CRKPVEKEILSNINGIMKPG-LNALIGPTGGGKSSLLDVLAARKDPSGL-SGDVLINGA 111
 Qy 138 PSSPQLVRKC-VAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELR 196
 | || :| | : : |||| | | :|| | : : : : | | ||
 Db 112 PRPANF--KCN SGYVVQDDVVMGTLTVRENLQFSAALRLATMTNHEKNERINRVIQELG 169
 Qy 197 LRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPTSGLD SFTAHLNVLKTL 256
 | : ||: || : ||: ||||: | ||: ||: : | || ||||: |||| ||: : |

Db 170 LDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLK 229
 Qy 257 RLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSN 316
 |::| |::|:|||| |::||:|:| |::| |::|:| |::| |::|
 Db 230 RMSKQGRTIIFSIIHQPRYSIFKLFDSTLLASGRMLMFHGAQAEALGYFESAGYHCEAYNN 289
 Qy 317 PADFYVDLTSIDRR-----SREQELATRE--KAQSLAALFLEKVRDL--DDFLWKAETK-- 366
 ||||::|:| |::|:| |::|:| |::|:| |::|:| |::|:| |::|:|
 Db 290 PADFFLDIINGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYK-ETKAE 348
 Qy 367 -----DLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLI 419
 : : | : : | : : | : : | : : | : : | : : | : : |
 Db 349 LHQLSGGEKKKKITVFKEISYTTSC-----HQLRWVSKRSFKNLLGNPQASIA 397
 Qy 420 HGAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISCYS----- 473
 : : | | : || : : : : | : || : : : : | : || : : : :
 Db 398 QIIVTVVLGLVIGAIYFGLKNDSTGIQNRAGVLFFL-----TTNQCFSSVSAVE 446
 Qy 474 ----ERAMLYYELEDGLYTTGPYFFAKILGE-LPEHCAYIIIIYGMPTYWLANLRPGLQPF 528
 |: : :| | | | | | | | | | | | | | | | | | | | | |
 Db 447 LFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLPMRMLPSIIFTCIVYFMLGLKPKADAF 506
 Qy 529 LLHFLLVWL VVFCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPA 588
 : : : | : | | | | | | | | | | | | | | | | | | | | |
 Db 507 FVMFTLMMVAYSASSMALAIAAGQSVSVATLLMTICFVFMFISGLLVNLTITIASWLS 566
 Qy 589 WISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT-----IAVSGDKIL--SAMELDSYP 639
 |: | | : | | : | : : | | | : : : | : : | : : | : :
 Db 567 WLQYFSIPRYGFTALQHNEFLGQNF-CPGLNATGNNPCNYATCTGEEYLVKQGIDLSPWG 625
 Qy 640 LYAIYLIVIGLSGGFMVLYYVSLRFIKQ 667
 |: : : : | : : | : | : : | : : | : : | : : | : :
 Db 626 LWKNHVALACMIVIFLTIAYLKLLFLKK 653

RESULT 14

US-10-405-806-2

; Sequence 2, Application US/10405806

; Publication No. US20030232362A1

; GENERAL INFORMATION:

; APPLICANT: KOMATANI, HIDEYA

; APPLICANT: HARA, YOSHIKAZU

; APPLICANT: KOTANI, HIDEHITO

; APPLICANT: NAKAGAWA, RINAKO

; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF

; FILE REFERENCE: 234985US0CONT

; CURRENT APPLICATION NUMBER: US/10/405,806

; CURRENT FILING DATE: 2003-04-03

; PRIOR APPLICATION NUMBER: PCT/JP01/08112

; PRIOR FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: JP2000-303441

; PRIOR FILING DATE: 2000-10-03

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 2

; LENGTH: 655

; TYPE: PRT

; ORGANISM: Homo sapiens
US-10-405-806-2

Query Match 18.3%; Score 642.5; DB 15; Length 655;
Best Local Similarity 27.2%; Pred. No. 2.4e-55;
Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

```
Qy      19 DTSG LQDR LFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFELAQFKMPWTSP 78
      :|:| | : : | | | | : : | : | | | : |
Db      16 NTNG-----FPATASNDLKAFTEGA--VLSFHNICYRVKLKSGF-----LP---- 54

Qy      79 SCQNSCELGI-QNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGGIKSGQIWINGQ 137
      |: | | | : : : | : | | : | | : : | | | | : | : | | |
Db      55 -CRKPVEKEILSNINGIMKPG-LNALIGPTGGGKSSLLDVLAARKDP SGL-SGDVLINGA 111

Qy     138 PSSPQLVRKC-VAHV RQHNL L PNLTVRET LAFIAQMRLPRTFSQAQRDKRVEDVIAELR 196
      | | | : | | : : : | | | | | | : | | | : : : | : | |
Db     112 PRPANF--KCSNGYVVQDDVVMGTLTVREN LQFS AALRLATMTNHEKNERINRVIQELG 169

Qy     197 LRQCADTRVG NMYVRGLSGGERRRV SIGVQLLWNP GILILDEPTSGLD SFTAHLVKTLS 256
      | : | : : | | : : | : | | | : | | | | : | | | | : : |
Db     170 LDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSS TANA VLLLLK 229

Qy     257 RLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSN 316
      | : : | | : : | : | | | | : | : | : | : | : | : | : |
Db     230 RMSKQGR TIIFS I HQPRYSIFKLFDSL TLLASGR LMFHGPAQEALGYFESAGYHCEAYNN 289

Qy     317 PADFYVDLTSIDRR---SREQELATRE--KAQSLAALFLEKVRDL--DDFLWKAETK-- 366
      | | | : : | : | : | : | : | : | : | : | : | : |
Db     290 PADFFLDIINGDSTAV ALNREEDFKATEIIEPSKQDKPLIEKLA EIYVNSSFYK-ETKAE 348

Qy     367 -----DLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLI 419
      : : | : : : | | : : | | : :
Db     349 LHQLSGGEKKKKITVFKEISYTT SFC-----HQLRWVSKRSFKNLLGNPQASIA 397

Qy     420 HGAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCY S----- 473
      : : : | | : | | : : : | : | : : : |
Db     398 QIIVTVVLGLVIGAIYFGLKNDSTGIQNRAGVLFFL-----TTNQCFSSVSAVE 446

Qy     474 ----ERAMLYYELEDGLYTTGPYFFAKILGE-LPEHCAYIIIIYGMPTYWLANLRPGLQPF 528
      | : : : | | | | | : | : | | : | : | : | : |
Db     447 LFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLPMRMLPSIIFTCIVYFMLGLKPKADAF 506

Qy     529 LLHFLLVWL VVFCRIMALAAAAALLPTFHMA SFFSNALYNSFYLAGGFMINLSSLWTVPA 588
      : : : | : | | | | | : : : : | : | : : : :
Db     507 FVMF TLMVAYSASSMALAIAAGQSVVSVATLLMTICFVFM MIFSGLLVNLT TIASWLS 566

Qy     589 WISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT-----IAVSGDKIL--SAMELDSYP 639
      | : | | : | | : | : | | | : : | : | : :
Db     567 WLQYFSIPRYGFTALQHNEFLGQNF-CPGLNATGNNPCNYATCTGEEYLVKQGDIDLS PWG 625

Qy     640 LYAIYLIVIGLSGGFMVLYYVSLRFIKQ 667
      | : : : : | : : | : | : :
Db     626 LWKNHVALACMIVIFLTIAYLKLLFLKK 653
```

US-09-961-086-1
; Sequence 1, Application US/09961086
; Publication No. US20030036645A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
; APPLICANT: ROSS, Douglas D.
; APPLICANT: DOYLE, L. Austin
; APPLICANT: ABRUZZO, Lynne
; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
; TITLE OF INVENTION: WHICH ENCODES IT
; FILE REFERENCE: EP19376-019
; CURRENT APPLICATION NUMBER: US/09/961,086
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/073,763
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/02577
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-086-1

Query Match 18.3%; Score 640.5; DB 10; Length 655;
Best Local Similarity 27.2%; Pred. No. 3.8e-55;
Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

Qy	19	DTSGQLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFELAQFKMPWTSP	78
		: : : : : : : :	
Db	16	NTNG-----FPATASNDLKAFTEGA--VLSFHNICYRVKLKSGF-----LP----	54
Qy	79	SCQNSCELGI-QNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQ	137
		: : : : : : : : : : : : :	
Db	55	-CRKPVEKEILSNINGIMKPG-LNALIGPTGGGKSSLLDVLAARKDPSPGL-SGDVLINGA	111
Qy	138	PSSPQLVRKC-VAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELR	196
		: : : : : : : :	
Db	112	PRPANF--KNSGYVVQDDVVMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIQELG	169
Qy	197	LRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLS	256
		: : : : : : : : : : : : : :	
Db	170	LDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLK	229
Qy	257	RLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSN	316
		: : : : : : : : : : : :	
Db	230	RMSKQGRTIIFSIIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQEALGYFESAGYHCEAYNN	289
Qy	317	PADFYVDLTSIDRR----SREQELATRE--KAQSLAALFLEKVRDL--DDFLWKAETK--	366
		: : : : : : : : : : : :	
Db	290	PADFFLDIINGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYK-ETKAE	348
Qy	367	-----DLDEDTCEVSSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLI	419
		: : : : : : : : : : : : : :	
Db	349	LHQLSGGEKKKITVFKEISYTSFC-----HQLRWVSKRSFKNLLGNPQASIA	397

Qy 420 HGAEACILMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKYS----- 473
 :: : || :||| : : | :|| : :||:|
 Db 398 QIIVTVVLGLVIGAIYFGLKNDSTGIQNRAGVLFFL-----TTNQCFSSVSAVE 446

 Qy 474 ----ERAMLYYELEDGLYTTGPYFFAKILGE-LPEHCAYIIIIYGMPYWLANLRPGLQPF 528
 |: : :| | | | | :| : | | | : | :| | :| |
 Db 447 LFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLPMTMLPSIIFTCIVYFMLGLKPKADAF 506

 Qy 529 LLHFLLVWLVVFCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPA 588
 : : :| : ||| | | :| : : : | :||::: : :
 Db 507 FVMMFTLMMVAYSASSMALAIAAGQSVVSVATLLMTICFVFMFIFSGLLVNLTTIASWLS 566

 Qy 589 WISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT-----IAVSGDKIL--SAMELDSYP 639
 |: | |: | | :| : : | | | :|:: | :| :
 Db 567 WLQYFSIPRYGFTALQHNEFLGQNF-CPGLNATGNNPCNYATCTGEEYLVKQGIDLSPWG 625

 Qy 640 LYAIYLIVIGLSGGFMVLYYVSLRFIKQ 667
 |: :: : : |: : |: | |:|:
 Db 626 LWKNHVALACMIVIFLTIAYLKLLFLKK 653

Search completed: February 27, 2004, 07:34:07
 Job time : 31.2443 secs

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:43 ; Search time 10.4203 Seconds
(without alignments)
3362.970 Million cell updates/sec

Title: US-09-989-981A-8
Perfect score: 3506
Sequence: 1 MAGKAAEERGLPKGATPQDT.....FMVLYYVSLRFIKQKPSQDW 673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	3502	99.9	673	1	ABG8_HUMAN	Q9h221 homo sapien
2	2873	81.9	673	1	ABG8_MOUSE	Q9dbm0 mus musculu
3	2814.5	80.3	694	1	ABG8_RAT	P58428 rattus norv
4	713	20.3	652	1	ABG5_RAT	Q99pe7 rattus norv
5	697	19.9	651	1	ABG5_HUMAN	Q9h222 homo sapien
6	691.5	19.7	652	1	ABG5_MOUSE	Q99pe8 mus musculu
7	656	18.7	687	1	WHIT_DROME	P10090 drosophila
8	653	18.6	1294	1	YOH5_YEAST	Q08234 saccharomyc
9	640.5	18.3	655	1	ABG2_HUMAN	Q9unq0 homo sapien
10	627	17.9	695	1	WHIT_ANOGA	Q27256 anopheles g
11	623.5	17.8	679	1	WHIT_CERCA	Q17320 ceratitidis c
12	621	17.7	666	1	ABG1_MOUSE	Q64343 mus musculu
13	620.5	17.7	677	1	WHIT_LUCCU	Q05360 lucilia cup
14	617	17.6	678	1	ABG1_HUMAN	P45844 homo sapien
15	600	17.1	598	1	YPC3_CAEEL	Q11180 caenorhabdi
16	583	16.6	709	1	WHIT_ANOAL	Q16928 anopheles a
17	573.5	16.4	646	1	ABG4_HUMAN	Q9h172 homo sapien

18	562.5	16.0	1049	1	ADP1_YEAST	P25371	saccharomyc
19	552	15.7	666	1	SCRT_DROME	P45843	drosophila
20	511	14.6	610	1	YQ5C_CAEEL	Q09466	caenorhabdi
21	464	13.2	1564	1	PDRA_YEAST	P51533	saccharomyc
22	463.5	13.2	675	1	BROW_DROME	P12428	drosophila
23	452.5	12.9	650	1	ABG3_MOUSE	Q99p81	mus musculu
24	437	12.5	668	1	BROW_DROVI	Q24739	drosophila
25	434.5	12.4	1499	1	CDR2_CANAL	P78595	candida alb
26	431	12.3	1529	1	PDRF_YEAST	Q04182	saccharomyc
27	424.5	12.1	1501	1	CDR1_CANAL	P43071	candida alb
28	412	11.8	1490	1	CDR4_CANAL	O74676	candida alb
29	401	11.4	1333	1	YN99_YEAST	P53756	saccharomyc
30	397	11.3	1501	1	SNQ2_YEAST	P32568	saccharomyc
31	388.5	11.1	1501	1	CDR3_CANAL	O42690	candida alb
32	388.5	11.1	1530	1	BFR1_SCHPO	P41820	schizosacch
33	385	11.0	1511	1	PDR5_YEAST	P33302	saccharomyc
34	349.5	10.0	1511	1	PDRC_YEAST	Q02785	saccharomyc
35	333	9.5	1410	1	PDRB_YEAST	P40550	saccharomyc
36	270.5	7.7	670	1	NRTC_SYNY3	P73450	synechocyst
37	252.5	7.2	894	1	YHIH_ECOLI	P37624	escherichia
38	251	7.2	371	1	Y4OS_RHISN	P55604	rhizobium s
39	250	7.1	1704	1	ABC3_HUMAN	Q99758	homo sapien
40	248.5	7.1	355	1	CYSA_SYNY3	P74548	synechocyst
41	248	7.1	362	1	AGLK_RHIME	Q9z3r9	rhizobium m
42	244	7.0	659	1	NRTC_SYNP7	P38045	synechococc
43	243	6.9	326	1	CYSA_PSESM	Q88as5	pseudomonas
44	241.5	6.9	332	1	SMOK_RHOSH	P54933	rhodobacter
45	241	6.9	236	1	CYSA_CHLVU	P56344	chlorella v

ALIGNMENTS

RESULT 1

ABG8_HUMAN

ID ABG8_HUMAN STANDARD; PRT; 673 AA.

AC Q9H221;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).

GN ABCG8.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA THR-231; GLN-263; ARG-574

RP AND ARG-596, AND VARIANT CYS-54.

RX MEDLINE=20553648; PubMed=11099417;

RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,

RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;

RT "Accumulation of dietary cholesterol in sitosterolemia caused by

RT mutations in adjacent ABC transporters.";

RL Science 290:1771-1775(2000).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), VARIANTS SITOSTEROLEMIA

RP HIS-184; THR-231; GLN-263; HIS-405; PRO-501; SER-543; PRO-572;
 RP GLU-574; ARG-574; ARG-596 AND PHE-570 DEL, AND VARIANTS HIS-19;
 RP CYS-54; LYS-238; VAL-259; LYS-400; ARG-575 AND ALA-632.
 RC TISSUE=Liver;
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 RA Patel S.B.;
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
 RT structure and spectrum of mutations involving sterolin-1 and
 RT sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
 RL Am. J. Hum. Genet. 69:278-290(2001).
 RN [3]
 RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
 CC ABCG5 along a pathway regulating dietary-sterol absorption and
 CC excretion.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9H221-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9H221-2; Sequence=VSP_000052;
 CC Note=Minor form detected in approximately 10% of the cDNA
 CC clones;
 CC -!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
 CC in the small intestine and colon. Detectable in a wide variety of
 CC human tissues.
 CC -!- DISEASE: Defects in ABCG8 are a cause of sitosterolemia
 CC [MIM:210250]; also known as phytosterolemia or shellfish
 CC sterolemia. It is a rare autosomal recessive disorder
 CC characterized by increased intestinal absorption of all sterols
 CC including cholesterol, plant and shellfish sterols, and decreased
 CC biliary excretion of dietary sterols into bile. Sitosterolemia
 CC patients have hypercholesterolemia, very high levels of plant
 CC sterols in the plasma, and frequently develop tendon and tuberous
 CC xanthomas, accelerated atherosclerosis and premature coronary
 CC artery disease.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.
 CC -!- CAUTION: Seems to have a defective ATP-binding region.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF320294; AAG40004.1; -.
 DR EMBL; AF324494; AAK84078.1; -.
 DR EMBL; AF351824; AAK84663.1; -.
 DR EMBL; AF351812; AAK84663.1; JOINED.
 DR EMBL; AF351813; AAK84663.1; JOINED.
 DR EMBL; AF351814; AAK84663.1; JOINED.
 DR EMBL; AF351815; AAK84663.1; JOINED.
 DR EMBL; AF351816; AAK84663.1; JOINED.
 DR EMBL; AF351817; AAK84663.1; JOINED.
 DR EMBL; AF351818; AAK84663.1; JOINED.
 DR EMBL; AF351819; AAK84663.1; JOINED.
 DR EMBL; AF351820; AAK84663.1; JOINED.
 DR EMBL; AF351821; AAK84663.1; JOINED.
 DR EMBL; AF351822; AAK84663.1; JOINED.
 DR EMBL; AF351823; AAK84663.1; JOINED.
 DR Genew; HGNC:13887; ABCG8.
 DR MIM; 605460; -.
 DR MIM; 210250; -.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Glycoprotein; Transmembrane; Transport; Alternative splicing;
 KW Polymorphism; Disease mutation.

FT	DOMAIN	1	416	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	417	437	1 (POTENTIAL).
FT	DOMAIN	438	447	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	448	468	2 (POTENTIAL).
FT	DOMAIN	469	492	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	493	513	3 (POTENTIAL).
FT	DOMAIN	514	531	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	532	552	4 (POTENTIAL).
FT	DOMAIN	553	569	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	570	590	5 (POTENTIAL).
FT	DOMAIN	591	639	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	640	660	6 (POTENTIAL).
FT	DOMAIN	661	673	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	619	619	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	376	376	Missing (in isoform 2).
FT				/FTId=VSP_000052.
FT	VARIANT	19	19	D -> H.
FT				/FTId=VAR_012250.
FT	VARIANT	54	54	Y -> C.
FT				/FTId=VAR_012251.
FT	VARIANT	184	184	R -> H (in sitosterolemia).
FT				/FTId=VAR_012252.
FT	VARIANT	231	231	P -> T (in sitosterolemia).
FT				/FTId=VAR_012253.
FT	VARIANT	238	238	E -> K.
FT				/FTId=VAR_012254.
FT	VARIANT	259	259	A -> V.
FT				/FTId=VAR_012255.
FT	VARIANT	263	263	R -> Q (in sitosterolemia).

FT				/FTId=VAR_012256.
FT	VARIANT	400	400	T -> K.
FT				/FTId=VAR_012257.
FT	VARIANT	405	405	R -> H (in sitosterolemia).
FT				/FTId=VAR_012258.
FT	VARIANT	501	501	L -> P (in sitosterolemia).
FT				/FTId=VAR_012259.
FT	VARIANT	543	543	R -> S (in sitosterolemia).
FT				/FTId=VAR_012260.
FT	VARIANT	570	570	Missing (in sitosterolemia).
FT				/FTId=VAR_012261.
FT	VARIANT	572	572	L -> P (in sitosterolemia).
FT				/FTId=VAR_012262.
FT	VARIANT	574	574	G -> E (in sitosterolemia).
FT				/FTId=VAR_012263.
FT	VARIANT	574	574	G -> R (in sitosterolemia).
FT				/FTId=VAR_012264.
FT	VARIANT	575	575	G -> R.
FT				/FTId=VAR_012265.
FT	VARIANT	596	596	L -> R (in sitosterolemia).
FT				/FTId=VAR_012266.
FT	VARIANT	632	632	V -> A.
FT				/FTId=VAR_012267.
SQ	SEQUENCE	673 AA;	75678 MW;	594AFD1D6C1BB50F CRC64;

Query Match 99.9%; Score 3502; DB 1; Length 673;
 Best Local Similarity 99.9%; Pred. No. 1.6e-251;
 Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLAS	60
Db	1	MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLAS	60
Qy	61	QVPWFEQLAQFKMPWTSPSCQNSCELGIONLSFKVRSGQMLAIIGSSGCGRASLLDVITG	120
Db	61	QVPWFEQLAQFKMPWTSPSCQNSCELGIONLSFKVRSGQMLAIIGSSGCGRASLLDVITG	120
Qy	121	RHGGGKIKSGQIWINGQPSSPQLVRKCAHVVRQHNQLLPNLTVRETAFIAQMRLPRTFS	180
Db	121	RHGGGKIKSGQIWINGQPSSPQLVRKCAHVVRQHNQLLPNLTVRETAFIAQMRLPRTFS	180
Qy	181	QAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPT	240
Db	181	QAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPT	240
Qy	241	SGLDSFTAHLVKTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM	300
Db	241	SGLDSFTAHLVKTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM	300
Qy	301	VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL	360
Db	301	VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL	360
Qy	361	WKAETKDLDEDTCEVSSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH	420
Db	361	WKAETKDLDEDTCEVSSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH	420

Qy 421 GAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY 480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 421 GAEACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY 480

Qy 481 ELEDGLYTTGPTYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWL VVF 540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 481 ELEDGLYTTGPTYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWL VVF 540

Qy 541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF 600
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF 600

Qy 601 EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYILIVIGLSGGFMVLYYV 660
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 601 EGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYILIVIGLSGGFMVLYYV 660

Qy 661 SLRFIKQKPSQDW 673
 ||||||||||||
 Db 661 SLRFIKQKPSQDW 673

RESULT 2

ABG8_MOUSE

ID ABG8_MOUSE STANDARD; PRT; 673 AA.
 AC Q9DBM0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
 GN ABCG8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 RA Patel S.B.;
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
 RT structure and spectrum of mutations involving sterolin-1 and
 RT sterolin-2, encoded by ABCG5 and ABCG8, respectively."
 RL Am. J. Hum. Genet. 69:278-290(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).

RN [3]

RP TISSUE SPECIFICITY, AND INDUCTION.

RX MEDLINE=20553648; PubMed=11099417;

RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,

RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;

RT "Accumulation of dietary cholesterol in sitosterolemia caused by
RT mutations in adjacent ABC transporters.";

RL Science 290:1771-1775(2000).

CC -!- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.

CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC ABCG5 along a pathway regulating dietary-sterol absorption and
CC excretion (By similarity).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1;

CC IsoId=Q9DBM0-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q9DBM0-2; Sequence=VSP_000053;

CC Note=No experimental confirmation available;

CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
CC level, in the liver.

CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
CC by the liver X receptor/retinoic acid X receptor (LXR/RXR) pathway.

CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC subfamily.

CC -!- CAUTION: Seems to have a defective ATP-binding region.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF324495; AAK84079.1; -.

DR EMBL; AK004871; BAB23630.1; -.

DR MGD; MGI:1914720; Abcg8.

DR InterPro; IPR003439; ABC_transporter.

DR Pfam; PF00005; ABC_tran; 1.

Qy	480	YELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVVLVV	539
		: :	
Db	480	YELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVVLVV	539
Qy	540	FCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWC	599
		: : : : :	
Db	540	FCCRTMALAASAMLPTFHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWC	599
Qy	600	FEGLMKIQFSRRYKMP LGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYY	659
		: : : : : : : : : :	
Db	600	FSGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDNLNSHPLYAIYLIVIGISYGFFLYY	659
Qy	660	VSLRFIKQKPSQDW	673
		: :	
Db	660	LSLKLKQKSIQDW	673

ABG8 RAT

```

ID      ABG8_RAT          STANDARD;          PRT;      694 AA.
AC      P58428; Q8CIQ5; Q923R7;
DT      28-FEB-2003 (Rel. 41, Created)
DT      15-MAR-2004 (Rel. 43, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
GN      ABCG8.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC      STRAIN=Sprague-Dawley;
RX      MEDLINE=21344600; PubMed=11452359;
RA      Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA      Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
RA      Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
RA      Patel S.B.;
RT      "Two genes that map to the STSL locus cause sitosterolemia: genomic
RT      structure and spectrum of mutations involving sterolin-1 and
RT      sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
RL      Am. J. Hum. Genet. 69:278-290(2001).
RN      [2]
RP      REVISIONS TO 3-4.
RA      Lu K., Yu H., Lee M.-H., Patel S.B.;
RL      Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND TISSUE SPECIFICITY.
RC      STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;
RC      TISSUE=Intestine, and Liver;
RX      PubMed=12783625;
RA      Yu H., Pandit B., Klett E., Lee M.-H., Lu K., Helou K., Ikeda I.,
RA      Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;
RT      "The rat STSL locus: characterization, chromosomal assignment, and
RT      genetic variations in sitosterolemic hypertensive rats.";
RL      BMC Cardiovasc. Disord. 3:4-4(2003).

```

CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
 CC ABCG5 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=3;
 CC IsoId=P58428-3; Sequence=Displayed;
 CC Name=1;
 CC IsoId=P58428-1; Sequence=VSP_008767;
 CC Name=2;
 CC IsoId=P58428-2; Sequence=VSP_008767, VSP_000054;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Highest expression in liver, with lower levels
 CC in small intestine and colon.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.
 CC -!- CAUTION: Seems to have a defective ATP-binding region.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF351785; AAK84831.2; -.
 DR EMBL; AY145899; AAN64276.1; -.
 DR EMBL; AF404109; AAK85393.1; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Glycoprotein; Transmembrane; Transport; Alternative splicing.
 FT DOMAIN 1 434 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 435 455 1 (POTENTIAL).
 FT DOMAIN 456 468 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 469 489 2 (POTENTIAL).
 FT DOMAIN 490 517 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 518 538 3 (POTENTIAL).
 FT DOMAIN 539 547 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 548 568 4 (POTENTIAL).
 FT DOMAIN 569 590 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 591 611 5 (POTENTIAL).
 FT DOMAIN 612 650 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 651 671 6 (POTENTIAL).
 FT DOMAIN 672 694 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 640 640 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 56 77 Missing (in isoform 1 and isoform 2).

FT /FTId=VSP_008767.
FT VARSPLIC 398 398 Missing (in isoform 2).
FT /FTId=VSP_000054.
FT CONFLICT 3 4 EK -> QT (IN REF. 3).
SQ SEQUENCE 694 AA; 78236 MW; 67F67C195F417587 CRC64;

Query Match 80.3%; Score 2814.5; DB 1; Length 694;
Best Local Similarity 77.4%; Pred. No. 1.3e-200;
Matches 538; Conservative 57; Mismatches 77; Indels 23; Gaps 2;

QY 1 MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQ----- 55
|| | || | | || | ||| :||||||||||||| ||||| ||
Db 1 MAEKTKEETQLWNGTVLQDASSLQDSVFSSES DNSLYFTYSGQSNLEVRDLTYQGGTCL 60

QY 56 -----VDLASQVPWFELQAQFKMPWTSPSCQNSCELGIQNLSFKVRSG 98
||:|||||||||||||:| | |:| :||:|||||||
Db 61 RSWGQEDPHMSLGLSESVDMA SQVPWFELQAQFKLPWRSRGSQDSWDLGIRNLSFKVRSG 120

QY 99 QMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSPQLVRKCVAVHVRQHNQLL 158
|||||||:|||||:||||||| |||:|||||||||:|:|:||||||| :|||
Db 121 QMLAIIGSAGCGRATLLDVITGRDHGGKMKSGQIWINQPPSTPQLIQKCVAVHVRQQDQLL 180

QY 159 PNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGER 218
||||||| |||||:||||| |||||:||||| |||||:|||||
Db 181 PNLTVRETLTFIAQMRLPKTFSQAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGER 240

QY 219 RRVSIGVQLLWNP GILILDEPTSGLD SFTA HNLVKTLSRLAKGNRLVLISLHQPRSDIFR 278
||||||| |||||:||||| |||||:||||| |||||:|||||
Db 241 RRVSIGVQLLWNP GILILDEPTSGLD SFTA HNLVRTLSRLAKGNRLVLISLHQPRSDIFR 300

QY 279 LFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELAT 338
||||||| |||||:||||| |||||:||||| |||||:|||||
Db 301 LFDLVLLMTSGTPIYLGVAQHMVQYFTSIGYPCPRYSNPADFYVDLTSIDRRSKEQEVAT 360

QY 339 REKAQSLAALFLEKVRDLDDFLWKAETKDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQ 398
|||: |||||: ||||| | | | ||| : :||: ||
Db 361 MEKARLLAALFLEKVQGFDDFLWKA EAKSLDTGT YAVSQTLTQDTNC-GTAAELPGMIQQ 419

QY 399 FTTLIRRQISNDFRDLPTLLIHGA EACLMSMTIGFLYFGHSGSIQLSFMDTAALLEFMIGAL 458
||||||| |||||: |||||: ||||| ||||| |||||
Db 420 FTTLIRRQISNDFRDLPTLFIHGA EACLMSLIIGFLYGHADKPLSFMDMAALLEFMIGAL 479

QY 459 IPFNVILDVISKCYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWL 518
|||||||:|||:|||:||||||| |||||:|||||||:||||| |||
Db 480 IPFNVILDVVS KCHSERSLLYYELEDGLYTAGPYFFAKVLGELPEHCAYVIIYGMPIYWL 539

QY 519 ANLRPGLQPFLHFLLVVLFCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMI 578
||||| : |||||:|:||||||| |||||:|:|||||:| ||||| |||||
Db 540 TNLRPGPELFLHFM LVLVVFCCRTMALAASAMLPTFHMSSFCCNALYNSFYLTAGFMI 599

QY 579 NLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTI AVSGDKILSAMELDSY 638
||:| |||||:||||||| |||:||| :||| :| || :||:|:|:
Db 600 NLNLWIVPAWISKMSFLRWCFSGLMQIQFNHGIYTTQIGNLTF SVPGDAMVTAMDLSH 659

QY 639 PLYAIYLIVIGLSGGFMVLYYVSLRFIKQKPSQDW 673
|||||||:| ||| :||:|:|||| |||
Db 660 PLYAIYLIVIGISCGFLSLYYLSLKFQKSIQDW 694

RESULT 4

ABG5_RAT

ID ABG5_RAT STANDARD; PRT; 652 AA.
AC Q99PE7; Q8CIQ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABCG5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Small intestine;
RX MEDLINE=20578753; PubMed=11138003;
RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
RA Dean M., Patel S.B.;
RT "Identification of a gene, ABCG5, important in the regulation of
RT dietary cholesterol absorption.";
RL Nat. Genet. 27:79-83(2001).
RN [2]
RP REVISION TO 2.
RA Lu K., Lee M.-H., Patel S.B.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND VARIANT CYS-583.
RC STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;
RX PubMed=12783625;
RA Yu H., Pandit B., Klett E., Lee M.H., Lu K., Helou K., Ikeda I.,
RA Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;
RT "The rat STSL locus: characterization, chromosomal assignment, and
RT genetic variations in sitosterolemic hypertensive rats.";
RL BMC Cardiovasc. Disord. 3:4-4(2003).
CC -!- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
CC ABCG8 along a pathway regulating dietary-sterol absorption and
CC excretion (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: Expressed only in liver and intestine.
CC -!- POLYMORPHISM: The polymorphism at position 583 is found in strains
CC SHR, SHRSP and Wistar Kyoto which are both hypertensive and
CC sitosterolemic. Strains which are hypertensive but not
CC sitosterolemic do not contain a polymorphism at this position.
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF312714; AAG53098.3; -.
DR EMBL; AY145899; AAN64275.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism.
FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 386 406 1 (POTENTIAL).
FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 423 443 2 (POTENTIAL).
FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 464 484 3 (POTENTIAL).
FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 505 525 4 (POTENTIAL).
FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 530 550 5 (POTENTIAL).
FT DOMAIN 551 624 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 625 645 6 (POTENTIAL).
FT DOMAIN 646 652 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 87 94 ATP (POTENTIAL).
FT CARBOHYD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 583 583 G -> C (in strains SHR, SHRSP and Wistar
FT Kyoto).
SQ SEQUENCE 652 AA; 73372 MW; 49FEF7372269299D CRC64;

Query Match 20.3%; Score 713; DB 1; Length 652;
Best Local Similarity 30.0%; Pred. No. 4.1e-45;
Matches 190; Conservative 115; Mismatches 232; Indels 96; Gaps 15;

QY 12 PKGAT-PQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQV-PWFEQLA 69
| : | | | : | | | : | : | | : : | | | : : : | | |
Db 9 PEGARGPHNNRGSQ----SSLEEGSV--TGSEARHSLGV--LNVSFVSNRVGPW----- 55

QY 70 QFKMPWTSPSCQNSCELGI-QNLSFKVRSGQMLAIIGSSGCGRASLLDVTGRGHGGKIK 128
| | | | : | : : | : | | : | : | | | : : | | | : | |
Db 56 -----WNIKSCQQKWDRKILKDVSLYIESGQTMCILGSSSGSKTTLDDAISGRLRRTGTL 110

QY 129 SGQIWINGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRV 188
| : : : | | : | : : | : | : | : | : | : | : | : | : | : |
Db 111 EGEVFVNGCELRRDQFQDCVSYLLQSDVFLSSLTVRETLRYTAMLAL-RSSSADFYDKKV 169

QY 189 EDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTFA 248
| | : | | | : | | : | | | | | | | | | : | : : | : | : | : |
Db 170 EAVLTELSLSHVADQMIGNYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTA 229

QY 249 HNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAG 308
: : : | | | : | : : : | : : : | : : : | : : : | : : : | : : : |

RT "Identification of a gene, ABCG5, important in the regulation of
RT dietary cholesterol absorption.";
RL Nat. Genet. 27:79-83(2001).
RN [3]
RP REVIEW.
RX MEDLINE=21474438; PubMed=11590207;
RA Schmitz G., Langmann T., Heimerl S.;
RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
RL J. Lipid Res. 42:1513-1520(2001).
RN [4]
RP VARIANTS SITOSTEROLEMIA GLN-146; HIS-389; PRO-419; HIS-419 AND
RP SER-550, AND VARIANT GLU-604.
RX MEDLINE=21344600; PubMed=11452359;
RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
RA Patel S.B.;
RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
RT structure and spectrum of mutations involving sterolin-1 and
RT sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
RL Am. J. Hum. Genet. 69:278-290(2001).
CC -!- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
CC ABCG8 along a pathway regulating dietary-sterol absorption and
CC excretion.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
CC in the small intestine and colon.
CC -!- DISEASE: Defects in ABCG5 are a cause of sitosterolemia
CC [MIM:210250]; also known as phytosterolemia or shellfish
CC sterolemia. It is a rare autosomal recessive disorder
CC characterized by increased intestinal absorption of all sterols
CC including cholesterol, plant and shellfish sterols, and decreased
CC biliary excretion of dietary sterols into bile. Sitosterolemia
CC patients have hypercholesterolemia, very high levels of plant
CC sterols in the plasma, and frequently develop tendon and tuberous
CC xanthomas, accelerated atherosclerosis and premature coronary
CC artery disease.
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF320293; AAG40003.1; -.
DR EMBL; AF312715; AAG53099.1; -.
DR Genew; HGNC:13886; ABCG5.
DR MIM; 605459; -.
DR MIM; 210250; -.

DR GO; GO:0030299; P:cholesterol absorption; NAS.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism;
 KW Disease mutation.
 FT DOMAIN 1 383 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 384 404 1 (POTENTIAL).
 FT DOMAIN 405 421 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 422 442 2 (POTENTIAL).
 FT DOMAIN 443 462 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 463 483 3 (POTENTIAL).
 FT DOMAIN 484 503 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 504 524 4 (POTENTIAL).
 FT DOMAIN 525 528 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 529 549 5 (POTENTIAL).
 FT DOMAIN 550 623 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 624 644 6 (POTENTIAL).
 FT DOMAIN 645 651 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 86 93 ATP (POTENTIAL).
 FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 146 146 E -> Q (in sitosterolemia).
 FT /FTid=VAR_012244.
 FT VARIANT 389 389 R -> H (in sitosterolemia).
 FT /FTid=VAR_012245.
 FT VARIANT 419 419 R -> H (in sitosterolemia).
 FT /FTid=VAR_012246.
 FT VARIANT 419 419 R -> P (in sitosterolemia).
 FT /FTid=VAR_012247.
 FT VARIANT 550 550 R -> S (in sitosterolemia).
 FT /FTid=VAR_012248.
 FT VARIANT 604 604 Q -> E.
 FT /FTid=VAR_012249.
 SQ SEQUENCE 651 AA; 72503 MW; 950BABFCBB6A1536 CRC64;

Query Match 19.9%; Score 697; DB 1; Length 651;
 Best Local Similarity 28.9%; Pred. No. 6.3e-44;
 Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

Qy 16 TPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFELAQFKMPW 75
 || : ||| | | : : || : : | : || : : : |
 Db 8 TPGGSMGLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQQW 64
 Qy 76 TSPSCQNSCELGIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI 134
 | : : : | | || : : |||| : : || : : | | : : :
 Db 65 TRQI-----LKDVSILYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTF-LGEVYV 115
 Qy 135 NGQPSSPQLVRKCVAVHRQHNQLLPNLTIVRETAFIAQMRLPRTFSQAQRDKRVEDVIAE 194
 || : : : | : : | : || : |||| : : | : : | : || : ||
 Db 116 NGRALRREQFQDCFSYVLQSDTLSSSLTVRETLYHTALLAI-RRGNPGSFQKKVEAVMAE 174
 Qy 195 LRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKT 254

```

      | |  || :|| :|:| ||||| ||| :| :: |||:| | ||: :|
Db      175 LSLSHVADRLIGNYSLGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGLDCMTANQIVVL 234
Qy      255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314
      | ||: ||:|:::|||||:::| || : :: | | : | | : :| |||| :
Db      235 LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEH 294
Qy      315 SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK 362
      ||| |||:||||:| :|:|:| : :| : : : :|::: |
Db      295 SNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHL----- 348
Qy      363 AETKDLDEDTCEVSSVTPLDTNCLPSPTK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHG 421
      |: :| || || : |:|| | |: ::
Db      349 -----KTLPM----VPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRL 390
Qy      422 AEACLSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSER 475
      : :| : : | |:|| | ||: |: :|: :: |
Db      391 LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DRVGLLYQFVGATPYTGMLNAVNLFPVLR 446
Qy      476 AMLYYELEDGLYTTGPFYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLV 535
      |: | :||| | | || :|: || | : |
Db      447 AVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLGHPEVAREF----- 499
Qy      536 WLVVFFCCRIMALAAAALLPTFHMASFFS-----NALYNSFYLAG-----GFM 577
      :||| : | : |:: :|| ||:
Db      500 -----GYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVALLSIAGVLVSGGFL 549
Qy      578 INLSSLWTPPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIASV 625
      |: : || :| :| | |: :| : |::: :
Db      550 RNIQEMPIPFKIIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTN 597

```

RESULT 6

ABG5_MOUSE

```

ID   ABG5_MOUSE          STANDARD;          PRT;          652 AA.
AC   Q99PE8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN   ABCG5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Liver;
RX   MEDLINE=20578753; PubMed=11138003;
RA   Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
RA   Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
RA   Dean M., Patel S.B.;
RT   "Identification of a gene, ABCG5, important in the regulation of
RT   dietary cholesterol absorption.";
RL   Nat. Genet. 27:79-83(2001).
RN   [2]
RP   TISSUE SPECIFICITY, AND INDUCTION.

```

RX MEDLINE=20553648; PubMed=11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by
 RT mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
 CC by the liver X receptor/retinoic X receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF312713; AAG53097.1; -.
 DR MGD; MGI:1351659; Abcg5.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 386 406 1 (POTENTIAL).
 FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 423 443 2 (POTENTIAL).
 FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 464 484 3 (POTENTIAL).
 FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 505 525 4 (POTENTIAL).
 FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 530 550 5 (POTENTIAL).
 FT DOMAIN 551 622 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 623 643 6 (POTENTIAL).
 FT DOMAIN 644 652 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 87 94 ATP (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;

Query Match 19.7%; Score 691.5; DB 1; Length 652;
Best Local Similarity 28.6%; Pred. No. 1.6e-43;
Matches 188; Conservative 129; Mismatches 241; Indels 99; Gaps 16;

Qy	45	NTLEVRDLNYQVDLASQV-PWFEQLAQFKMPWTSPSCQNSCELGI-QNLSFKVRSGQMLA	102
Db	37	HSLGVLHVSYSV--SNRVGPW-----WNIKSCQQKWDRQILKDVSLYIESGQIMC	84
Qy	103	IIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSSPQLVRKCAHVHRQHNQLLPNLT	162
Db	85	ILGSSSGSKTTLDDAISGRLLRRTGTLEGEVFVNGCELRRDQFQDCFSYVLQSDVFLSSLT	144
Qy	163	VRETAFIAQMRPLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVS	222
Db	145	VRETLRYTAMLALCRS-SADFYNNKKVEAVMTELSLSHVADQMIGSYNFGGISSGERRRVS	203
Qy	223	IGVQLLWNPGILILDEPTSGLDSTAHNLVKTLSRLAKGNRLVLISLHQPRSDIFRLFDL	282
Db	204	IAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLLAELARRDRIVITIHQPRSELFQHFDD	263
Qy	283	VLLMTSGTPIYLGAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKA	342
Db	264	IAILTYGELVFCGTPPEMLGFFNNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRV	323
Qy	343	QSLAALFLEKVRDLDDFLWKAETKDLDEDTCESSVTPLDTNCLPSPTK-MPGAVQQFTT	401
Db	324	QMLECAFKE-----SDIYHKI-LENIERARYLKTLP-----VPFKTKDPPGMFGKLG	371
Qy	402	LIRRQISNDFRDLPTLLIHGAACILMSMTIGF--LYFGHGSIQLSFMDTAALLFMIGALI	459
Db	372	LLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDRVGLLYQLVGAT	431
Qy	460	PFNVILDVISKCYSERAMLYYELEDGLYTTGYPYFFAKILGELPEHCAYIIYGMPTYWLA	519
Db	432	PYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIATVIFSSVCYWT	491
Qy	520	NLRPGLQPFLHFLVLVVFCCRIMALAAAALLPTFHMASFFSNAL-----	566
Db	492	GLYPEVARF-----GYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSI	534
Qy	567	-----YNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFEGLMKIQFSRRTYKMPGLNLT	621
Db	535	VALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEF----YGL---NFT	587
Qy	622	IAVSGDKILSAMELDSYPLYAI-----YLVIGLSSGGFMVL	657
Db	588	CGGSNTSML-----NHPMCAITQGVQFIEKTCPGATSRFTANFLILYGFIPALVIL	638

RESULT 7

WHIT DROME

ID WHIT DROME STANDARD; PRT; 687 AA.

AC P10090; Q9V3A2; Q9XY33;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE White protein.
 GN W OR EG:BACN33B1.1 OR CG2759.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Head;
 RX MEDLINE=90221897; PubMed=2109311;
 RA Pepling M., Mount S.M.;
 RT "Sequence of a cDNA from the *Drosophila melanogaster* white gene."
 RL Nucleic Acids Res. 18:1633-1633(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85134865; PubMed=6084717;
 RA O'Hare K., Murphy C., Levis R., Rubin G.M.;
 RT "DNA sequence of the white locus of *Drosophila melanogaster*."
 RL J. Mol. Biol. 180:437-455(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21100348; PubMed=11156992;
 RA Lukacsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niwa S.,
 RA Yamamoto D.;
 RT "Dual-tagging gene trap of novel genes in *Drosophila melanogaster*."
 RL Genetics 157:727-742(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=Oregon-R;

RX MEDLINE=20196011; PubMed=10731137;

RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,
 RA Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.;

RT "From sequence to chromosome: the tip of the X chromosome of *D.*
 RT *melanogaster*.";

RL Science 287:2220-2222(2000).

RN [6]

RP SEQUENCE OF 224-331 FROM N.A.

RX MEDLINE=89339145; PubMed=2503416;

RA Tearle R.G., Belote J.M., McKeown M., Baker B.S., Howells A.J.;

RT "Cloning and characterization of the scarlet gene of *Drosophila*
 RT *melanogaster*.";

RL Genetics 122:595-606(1989).

CC -!- FUNCTION: Part of a membrane-spanning permease system necessary
 CC for the transport of pigment precursors into pigment cells
 CC responsible for eye color. White dimerize with brown for the
 CC transport of guanine and with scarlet for the transport of
 CC tryptophan.

CC -!- SUBUNIT: Heterodimer of white with either brown or scarlet.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X51749; CAA36038.1; -.

DR EMBL; X02974; CAA26716.1; -.

DR EMBL; AB028139; BAA78210.1; -.

DR EMBL; AE003425; AAF45826.1; -.

```

DR      EMBL; AL133506; CAB65847.1; -.
DR      EMBL; X76202; CAA53795.1; -.
DR      PIR; S08635; FYFFW.
DR      FlyBase; FBgn0003996; w.
DR      GO; GO:0004888; F:transmembrane receptor activity; NAS.
DR      GO; GO:0006727; P:ommochrome biosynthesis; IMP.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      InterPro; IPR005284; Pigment_permease.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.
DR      TIGRFAMS; TIGR00955; 3a01204; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW      Pigment; ATP-binding; Transmembrane; Transport.
FT      NP_BIND      130      137      ATP (BY SIMILARITY).
FT      TRANSMEM     435      453      POTENTIAL.
FT      TRANSMEM     465      485      POTENTIAL.
FT      TRANSMEM     515      533      POTENTIAL.
FT      TRANSMEM     542      563      POTENTIAL.
FT      TRANSMEM     576      594      POTENTIAL.
FT      TRANSMEM     659      678      POTENTIAL.
FT      CONFLICT      25      29      GDSGA -> LIFEIPYHCRVTAD (IN REF. 2 AND
FT                                     3).
FT      CONFLICT      49      49      L -> R (IN REF. 4 AND 5).
FT      CONFLICT     335     371      VGAQCPTNYPADFYVQVLAVVPGREIESRDRIAKIC ->
FT      ITLHLNSYPAWVPSVLPTTIRRTFTYRCWPLCPDGRSSPVI
FT      GSPRYG (IN REF. 3).
SO      SEQUENCE     687 AA; 75672 MW; 24AFAD799DE0D396 CRC64;

```

[illegible]

DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 2.
 DR ProDom; PD000006; ABC_transporter; 2.
 DR SMART; SM00382; AAA; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
 KW Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;
 KW Transport; Repeat.
 FT TRANSMEM 376 396 POTENTIAL.
 FT TRANSMEM 496 516 POTENTIAL.
 FT TRANSMEM 531 551 POTENTIAL.
 FT TRANSMEM 605 625 POTENTIAL.
 FT TRANSMEM 1039 1059 POTENTIAL.
 FT TRANSMEM 1121 1141 POTENTIAL.
 FT TRANSMEM 1267 1287 POTENTIAL.
 FT NP_BIND 62 69 ATP (POTENTIAL).
 FT NP_BIND 727 734 ATP (POTENTIAL).
 FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 983 983 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1294 AA; 145157 MW; C555500A45E9284E CRC64;

Query Match 18.6%; Score 653; DB 1; Length 1294;
 Best Local Similarity 30.1%; Pred. No. 2.7e-40;
 Matches 171; Conservative 111; Mismatches 239; Indels 48; Gaps 13;

Qy 88 IQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQI----- 132
 : | : || :|::| || |: :||::| : || :| |
 Db 45 VNTFSMDLPSGSVMAMVGGSGSKTTLLNLVASKISGGLTHNGSIRYVLEDTGSEPNETE 104
 Qy 133 ----WINGQPSSPQLVRKCAVAVHRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKR- 187
 :|| | :||: | : | | || | | :| : :| | :
 Db 105 PKRAHLDGQ-DHPIQKHVIMAYLPQQDVLSPRLTCRETLKFAADLKL----NSSERTKKL 159
 Qy 188 -VEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSE 246
 || :| || |: |||| ||: |||||::||:| ||: || |: ||||:||||:
 Db 160 MVEQLIEELGLKDCADTLVGDNSHRGLSGGEKRRLSIGTQMISNPSIMFLDEPTTGLDAY 219
 Qy 247 TAHNLVKTL SRLAK-GNRLVLISLHQPRSDIFRLFDLVLLMTSGTPPIYLGAQHVMVQYFT 305
 :| :||| :|| | :|:|||||| | | |::| :| : : ||
 Db 220 SAFLVIKTLKKLAKEDGRTFIMSIHQPRSDILFLLDQVCILSKGNVVYCDKMDNTIPYFE 279
 Qy 306 AIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFLWKAET 365
 :||| |: ||||:|:|:| || :||: | | : : | : :
 Db 280 SIGYHVPQLVNPADYFIDLSSVDSRSKDKEEAATQSRNLNSL----IDHWHD-----YERTH 330
 Qy 366 KDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRQISNDFRDLPTLLIHGAEC 425
 | : : |: | : : |::| :| | | | :| | ||: ||
 Db 331 LQLQAESYI-SNATEIQIQNM--TTRLP-FWKQVTVLTRNFKLNFSDYVTLISTFAEPL 386

Qy 426 LMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIP--FNVILDVISKCYSERAMLYYELE 483
 :: |::|: : : | :: : | : | : |
 Db 387 IIGTVCGWIYYKPKDKSSIGGLRTTTACLYASTILQCYLYLLFDTYRLCEQDIALYDRERA 446
 Qy 484 DGLYTTGPFYFFA-KILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWL VVFC 542
 :| | : | || | : | :|: ||:: | : | | :|:| |
 Db 447 EGSVTPLAFIVARKISLFLSDDFAMTMIFVSITYFMFGLEADARKFFYQFAVVFLCQLSC 506
 Qy 543 RIMALAAAALLPTFHMAFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCFEG 602
 :: : |: | || | : : ||:| : || :| : |
 Db 507 SGLSMLSVAVSRDFSKASLVGNMTFTVLSMGCFFVNKVMVPVYVRWIKYIAFTWYSFGT 566
 Qy 603 LMKIQFSRR-----TYKMPLGNLTI AVSG 626
 || |: ||| : ||
 Db 567 LMSSTFTNSYCTTDNLDECLGNQILEVYG 595

RESULT 9

ABG2_HUMAN

ID ABG2_HUMAN STANDARD; PRT; 655 AA.
 AC Q9UNQ0; O95374; Q9BY73; Q9NUS0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-
 DE binding cassette transporter) (Breast cancer resistance protein).
 GN ABCG2 OR ABCP OR BCRP OR BCRP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=99065313; PubMed=9850061;
 RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;
 RT "A human placenta-specific ATP-binding cassette gene (ABCP) on
 RT chromosome 4q22 that is involved in multidrug resistance."
 RL Cancer Res. 58:5337-5339(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast cancer;
 RX MEDLINE=99080071; PubMed=9861027;
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RA Ross D.D.;
 RT "A multidrug resistance transporter from human MCF-7 breast cancer
 RT cells."
 RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).
 RN [3]
 RP ERRATUM.
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RA Ross D.D.;
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Kage K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.,

RA Sugimoto Y.;
 RT "Breast cancer resistance protein constitutes a 140-kDa complex as a
 RT homodimer.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
 RN [5]
 RP SEQUENCE OF 198-655 FROM N.A.
 RC TISSUE=Placenta;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
 RA Ninomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
 RN [6]
 RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Xenobiotic transporter that appears to play a major role
 CC in the multidrug resistance phenotype of a specific MCF-7 breast
 CC cancer cell line. When overexpressed, the transfected cells become
 CC resistant to mitoxantrone, daunorubicin and doxorubicin, display
 CC diminished intracellular accumulation of daunorubicin, and
 CC manifest an ATP-dependent increase in the efflux of rhodamine 123.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF103796; AAD09188.1; -.
 DR EMBL; AF098951; AAC97367.1; -.
 DR EMBL; AB056867; BAB39212.1; -.
 DR EMBL; AK002040; BAA92050.1; -.
 DR Genew; HGNC:74; ABCG2.
 DR MIM; 603756; -.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR GO; GO:0005524; F:ATP binding; TAS.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; TAS.
 DR GO; GO:0005215; F:transporter activity; TAS.
 DR GO; GO:0008559; F:xenobiotic-transporting ATPase activity; TAS.
 DR GO; GO:0009315; P:drug resistance; TAS.
 DR GO; GO:0006810; P:transport; TAS.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.

```

DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transmembrane; Transport.
FT DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 396 416 POTENTIAL.
FT DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 429 449 POTENTIAL.
FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 478 498 POTENTIAL.
FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 507 527 POTENTIAL.
FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 536 556 POTENTIAL.
FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 631 651 POTENTIAL.
FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 80 87 ATP (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (PO
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (PO
FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (PO
FT CONFLICT 24 24 V -> A (IN REF. 2 AND 4).
FT CONFLICT 166 166 E -> Q (IN REF. 2 AND 4).
FT CONFLICT 208 208 F -> S (IN REF. 1).
FT CONFLICT 315 316 MISSING (IN REF. 5).
FT CONFLICT 482 482 R -> T (IN REF. 2).
SO SEQUENCE 655 AA; 72343 MW; 89A6D3511DC5CCE0 CRC64;

```

Qy	80	CQNSCELGIG-QNLSFKVRSQMLAIIGSSGCGRASLLDVTITGRGHGKIKSGQIWINQGP	138
Db	55	CRKPVEKEILSNINGIMKPG-LNAILGPTGGGKSSLLDVLAAARKDPSGL-SGDVLINGAP	112
Qy	139	SSPQLVRKC-VAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAELRL	197
Db	113	RPANF--KCNSGYVQDDVVMGTLTVRENLFSAALRLATTMTNHEKNERINRVIEELGL	170
Qy	198	RQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHNLVKTLRS	257
Db	171	DKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKR	230
Qy	258	LAKGNRLVLISLHQPRSDIFRLEFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNP	317
Db	231	MSKQGRTIIFSIIHQPRYSIFKLFDLSLTLLASGRLMFHGPAQEALGYFESAGYHCEAYNNP	290
Qy	318	ADFYVDLTSIDRR----SREQELATRE--KAQSLAALFLEKVRDL--DDFLWKAETK---	366
Db	291	ADFFLDIINGDSTAVALNREEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYK-ETKAEL	349
Qy	367	-----DLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH	420
Db	350	HQLSGGEKKKKITVFKEISYTTSEF-----HQLRWVSKRSFKNLLGNPQASIAQ	398
Qy	421	GAEACLSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKYS-----	473

```

      :: : || :||| :      : | :|| :      ::|||
Db      399 IIVTVVLGLVIGAIYFGLKNDSTGIQNRAGVLFLL-----TTNQCFSSVSAVEL 447

Qy      474 ---ERAMLYYELEDGLYTTGPYFFAKILGE-LPEHCAYIIIIYGMPTYWLANLRPGLQPFLL 529
      | : : :| | | | | :| : | | | | :| :| :| |
Db      448 FVVEKKLFIHEYISGYRVSSYFLGKLLSDLLPMRMLPSIIFTCIVYFMLGLKPKADAFF 507

Qy      530 LHFLLVVLVVFCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAW 589
      : : :| : | ||| | | :| : : : | :| :| : : :|
Db      508 VMMFTLMMVAYSASSMALAIAAGQSVSVATLLMTICFVFMIFSGLLVNLTTIASWLSW 567

Qy      590 ISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT-----IAVSGDKIL--SAMELDSYPL 640
      : | | :| | | :| : : | | | :| :| :| :| :|
Db      568 LQYFSIPRYGFTALQHNEFLGQNF-CPGLNATGNNPCNYATCTGEEYLVKQGIDLSPWGL 626

Qy      641 YAIYLVIVIGLSGGFMVLYYVSLRFIKQ 667
      : : : : | : : | : | :| :|
Db      627 WKNHVALACMIVIFLTIAYLKLFLK 653

```

RESULT 10

WHIT_ANOGA

```

ID   WHIT_ANOGA          STANDARD;          PRT;          695 AA.
AC   Q27256; Q17006;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   White protein.
GN   W.
OS   Anopheles gambiae (African malaria mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=7165;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Suakoko / G3;
RX   MEDLINE=96423158; PubMed=8825759;
RA   Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hilfiker D.,
RA   Collins F.H.;
RT   "Cloning and characterization of the white gene from Anopheles
RT   gambiae.";
RL   Insect Mol. Biol. 4:217-231(1995).
CC   -!- FUNCTION: May be part of a membrane-spanning permease system
CC   necessary for the transport of pigment precursors into pigment
CC   cells responsible for eye color.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U29486; AAC46995.1; -.

```



```

Db      363 -----EAECRDMIKKICDSFAVSPIAREVLETASVAGKG 396
Qy      391 -KMPGAVQ-----QFTTLIRRQISNDFRDLPTLLIHGAEACILMSMTIGF 433
      | :|          || :: | : :| : : : :| |
Db      397 MDEPYMLQQVEGVGSTGYRSSWWTQFYCILWRSWLSVLKDPMLVKVRLLOTAMVATLIGS 456
Qy      434 LYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYYELEDGLYTTGPYF 493
      :||| | | : || : | : || : :| : | || ||
Db      457 IYFGQVLDQDGVMNINGSLSLFLTNMTFQNVFAVINVFSaelPVFLREKRSRLYRVDTYF 516
Qy      494 FAKILGELPEHCAYIIIIYGMPYWLANLRPGLQPFLHFLLVWLVVFCRIMALAAAALL 553
      | : ||| | : : || : || | :| :| || :
Db      517 LGKTIAELPLFIAPVPFVFTSITYPMIGLRTGATHYLTTLFIVTLVANVSTSFGYLISCAS 576
Qy      554 PTFHMASFFSNALYNSFYLAGGFMINLSSLWTPPAWISKVSFLRW---CFEGLMKIQFS- 609
      : || : | : ||| :| :| ||| : :| :| | :| :|
Db      577 SSISMALSVGPPVVIPFLIFGGFFLNSAS---VPAYFKYLSYLSWFRYANEALLINQWST 633
Qy      610 -----RRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLY 658
      | | : : : : || | :| ||| : | :|
Db      634 VVDGEIACTRANVTCPRSEIILETFNFRV-EDFALDIACLEFA--LIVLFRLGALLCLW 688

```

RESULT 11

WHIT_CERCA

ID WHIT_CERCA STANDARD; PRT; 679 AA.

AC Q17320;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE White protein.

GN W.

OS Ceratitis capitata (Mediterranean fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Tephritoidea; Tephritidae; Ceratitis.

OX NCBI_TaxID=7213;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96123276; PubMed=8533095;

RA Zwiebel L.J., Saccone G., Zacharopoulou A., Besansky N.J.,

RA Favia G., Collins F.H., Louis C., Kafatos F.C.;

RT "The white gene of Ceratitis capitata: a phenotypic marker for
germline transformation.";

RL Science 270:2005-2007(1995).

CC -!- FUNCTION: May be part of a membrane-spanning permease system
necessary for the transport of pigment precursors into pigment
cells responsible for eye color.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X89933; CAA61998.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR005284; Pigment_permease.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 121 128 ATP (BY SIMILARITY).
FT TRANSMEM 427 445 POTENTIAL.
FT TRANSMEM 457 477 POTENTIAL.
FT TRANSMEM 507 525 POTENTIAL.
FT TRANSMEM 534 555 POTENTIAL.
FT TRANSMEM 568 586 POTENTIAL.
FT TRANSMEM 651 670 POTENTIAL.
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 679 AA; 75145 MW; 3F9CBC78A835C4CC CRC64;

Query Match 17.8%; Score 623.5; DB 1; Length 679;
Best Local Similarity 28.3%; Pred. No. 1.8e-38;
Matches 169; Conservative 112; Mismatches 264; Indels 53; Gaps 8;

Qy 88 IQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGG-KIKSGQI-WINGQPSSPQLVR 145
::| | |::||:|||| |: ||: | | :| | : ::
Db 104 LKNDSGVAYPGELLAVMGSSGAGKTTLLNASAFRSSKGVQISPSTIRMLNGHPVDAKEMQ 163
Qy 146 KCVAVHRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRV 205
|:|:| : : :|| || | | :|:| :| :| :|:| || :| | :| :| :
Db 164 ARCAYVQQDDLFIGSLTAREHLIFQAMVRMPRHMTQKQKVQRVDQVIQDLSLGKCQNTLI 223
Qy 206 G-NMYVRGLSGGERRRVISIGVQLLWNPGLILDEPTSGLDSFTAHNVLKTL SRLAKGNRL 264
| :|||||||:|: : | :| :|| ||||| |||:|: | :|: :
Db 224 GVPGRVKGLSGGERKRLAFASEALTDPPLLICDEPTSGLDSFMAHSVQVLKCLSQKGKT 283
Qy 265 VLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDL 324
|:::| || |:| ||| :|| | :|| | :|: || || ||||| :
Db 284 VILTIHQPSSELFELFDKILLMAEGRVAFLGTPGEAVDFFSYIGATCPTNYTPADFYVQV 343
Qy 325 TS-----IDRRSREQEL-----ATREKAQSLAALFLEK---VRDLDDFLWKAETKD 367
: : :| | : : :|| |: | | : : :||
Db 344 LAVVPGREVESRDRVAKICDNFAVGKVSREMEQNFQKLVKSNGFGKEDENEYTYKASW-- 401
Qy 368 LDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGAECIM 427
|| : : | : : : : : : :
Db 402 -----FMQFRAVLWRSWLSVLKEPLLVKVRLQLTTMV 433
Qy 428 SMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYYELEDGLY 487
: : || : : | : | : :| : | : : | : | ||
Db 434 AVLIGLIFLGQQLTQVGVMNINGAIFLFLTNMTFQNSFATITVFTTELPVFMRETRSRLY 493
Qy 488 TTGPYFFAKILGELPEHCAYIIIIYGMPYWLANLRPGLQPFLHFLLVLVVFCCRIMAL 547


```

      || | : |||      :: | | |||: | || ||
Db      494 RCDTYFLGKTIAELPLFLVVPFLFTAIAAYPLIGLRPGVDHFFTALALVTLVANVSTSFGY 553

Qy      548 AAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFEGLMKIQ 607
      : : || : | | ||| :| | : | :| :| : ||| : |
Db      554 LISCACSSTSMALSVGPPVIIPFLFEGGFFLNSGSVPVYFKWLSYLSWFRYANEGLLINQ 613

Qy      608 FS---RRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLVIGLSGGFMVLYYVSL 662
      :: | | | ||: || : : | :: : | | : |::|
Db      614 WADVKPGEITCTLSNTTCPSSGEVILETLNFSASDLPFDFIGLALLIVGFRISAYIAL 671

```

RESULT 12

ABG1_MOUSE

```

ID  ABG1_MOUSE      STANDARD;      PRT;      666 AA.
AC  Q64343;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  15-MAR-2004 (Rel. 43, Last annotation update)
DE  ATP-binding cassette, sub-family G, member 1 (White protein homolog)
DE  (ATP-binding cassette transporter 8).
GN  ABCG1 OR ABC8 OR WHT1.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=97186700; PubMed=9034316;
RA  Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
RA  Goldenson D., Son D., Arciniegas S., Wu R.;
RT  "Isolation and characterization of a mammalian homolog of the
RT  Drosophila white gene.";
RL  Gene 185:77-85(1997).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=DBA/2;
RX  MEDLINE=96359154; PubMed=8703120;
RA  Savary S., Denizot F., Luciani M.-F., Mattei M.-G., Chimini G.;
RT  "Molecular cloning of a mammalian ABC transporter homologous to
RT  Drosophila white gene.";
RL  Mamm. Genome 7:673-676(1996).
RN  [3]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21092576; PubMed=11162488;
RA  Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
RA  Assmann G., Cullen P.;
RT  "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
RL  Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN  [4]
RP  INDUCTION, AND PROBABLE FUNCTION.
RX  MEDLINE=20261604; PubMed=10799558;
RA  Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
RA  Mangelsdorf D.J., Edwards P.A.;
RT  "Human white/murine ABC8 mRNA levels are highly induced in
RT  lipid-loaded macrophages. A transcriptional role for specific
RT  oxysterols.";

```

RL J. Biol. Chem. 275:14700-14707(2000).
 RN [5]
 RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
 CC an active component of the macrophage lipid export complex. Could
 CC also be involved in intracellular lipid transport processes. The
 CC role in cellular lipid homeostasis may not be limited to
 CC macrophages.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- TISSUE SPECIFICITY: Expressed mainly in brain, thymus, lung,
 CC adrenals, spleen and placenta. Little or no expression in liver,
 CC kidney, heart, muscle or testes.
 CC -!- INDUCTION: Strongly induced in macrophage cell line RAW264.7
 CC during cholesterol influx. Induction is mediated by the liver X
 CC receptor/retinoide X receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U34920; AAB47738.1; -.
 DR EMBL; Z48745; CAA88636.1; -.
 DR EMBL; AF323659; AAK27442.1; -.
 DR MGD; MGI:107704; Abcg1.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF000005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMS; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Transport; Lipid transport; ATP-binding; Transmembrane.
 FT DOMAIN 1 414 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 415 433 POTENTIAL.
 FT DOMAIN 434 444 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 445 465 POTENTIAL.
 FT DOMAIN 466 494 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 495 513 POTENTIAL.
 FT DOMAIN 514 521 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 522 543 POTENTIAL.
 FT DOMAIN 544 555 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 556 574 POTENTIAL.
 FT DOMAIN 575 637 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 638 657 POTENTIAL.
 FT DOMAIN 658 666 CYTOPLASMIC (POTENTIAL).

FT NP_BIND 118 125 ATP (POTENTIAL).
SQ SEQUENCE 666 AA; 74033 MW; EDDC6AFBD43950B6 CRC64;

Query Match 17.7%; Score 621; DB 1; Length 666;
Best Local Similarity 25.4%; Pred. No. 2.7e-38;
Matches 165; Conservative 137; Mismatches 279; Indels 68; Gaps 16;

```
Qy      33 DNSLYFT----YSGQPN----TLEVRDLNYQVDLASQVPWFQQLAQFKMPWTSPSCQNSC 84
      ||: || :| | :| :||:| | : ||:: :
Db      57 DNN--FTEAQRFSSLPRRAAVNIEFKDLSYSV---PEGPWKKKKGYKTL----- 100

Qy      85 ELGIQNLSEFKVRSQMLAIIGSSGCGRASLLDVTGRGHGGKIKSGQIWIWQSPSSPOLV 144
      :: :| | ||::||:| || ||::|::: | | | : ||| | :
Db     101 ---LKGISGKFNSGELVAIMGPSGAGKSTLMNILAGYRETG--MKGAVLINGMPRDLRCF 155

Qy     145 RKCVAHVRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTR 204
      || :: | : |||:| | : | :| | : |::: | | ||:| |
Db     156 RKVSCYIMQDDMLLPHLTVQEAMMVSAAHLKLQE--KDEGRREMVKEILTALGLLPCANTR 213

Qy     205 VGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLSRLAKGNRL 264
      |: |||:|:|:| :|:| || :: ||||| | : | : ||:| |
Db     214 TGS-----LSGGQRKRLAIALELVNPPVMFFDEPTSGLDASCFQVVSIMKGLAQGGRS 268

Qy     265 VLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDL 324
      :: :||| : :| ||| : :: | :| | :| | :| || | |||| ::
Db     269 IVCTIHQPSAKLFELFDQLYVLSQGCQVYRGKVSNLVPYLRDLGLNCPTYHNPADFVMEV 328

Qy     325 TSIDRRSREQELATREKAQSLAALFLEKV---RDLDLFLWKAETKDLDEDTCVESSVTPL 381
      | : : | : | : : ||| : :||:
Db     329 ASGEYGDQNSRLVRVAVREGMCDADYKRDLDGGTDVNPFLWH---RPAEEDSASMEGCHSF 385

Qy     382 DTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGAELMSMTIGFLYFGHSGI 441
      :|| || | :| : || : : || || | | :
Db     386 SASCL-----TQFCILFKRTFLSIMRDSVLTHLRITSHIGIGLLIGLLYLIGNE 435

Qy     442 QLSFMDTAALLFMIGALIPFNVLVDISKYSERAMLYYELEDGLYTTGPYFFAKILGEL 501
      : : || : | : : | : : | : | : || : :
Db     436 AKKVLNSGFLFFSMLFLMFAALMPTVLTFPLEMSVFLREHLNYWYSLKAYYLAKTMADV 495

Qy     502 PEHCAYIIIIYGMPTYWLANLRPGLQPFLHLLVWLNVFCCRIMALAAAALLPTFHMASF 561
      | : : | ||: : | : | : : | | : :|:|
Db     496 PFQIMFPVAYCSIVYWMTSQPSDAVRFLFAALGTMTSLVAQSLGLLIGAASTSLQVATF 555

Qy     562 FSNALYNSFYLAGGFMINLSSLWTVPA---WISKVSFLRWCFEGLMKIQF--SRRTYKMP 616
      | || : : | :| | :| :|:| : ||: : |
Db     556 VGPVTAIPVLLFSGFFVSFD---TIPAYLQWMSYISYVRYGFEGVILSIYGLDREDLHCD 612

Qy     617 LGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYVSLRFI 665
      : : || :::: || : ||:| :::||| |
Db     613 IAETCHFQKSEAILRELDVENAKLY-LDFIVLG-----IFFISLRLI 653
```

RESULT 13

WHIT_LUCCU

ID WHIT_LUCCU STANDARD; PRT; 677 AA.

AC Q05360;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN W.
 OS *Lucilia cuprina* (Greenbottle fly) (Australian sheep blowfly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;
 OC Calliphoridae; *Lucilia*.
 OX NCBI_TaxID=7375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97087158; PubMed=8933176;
 RA Garcia R.L., Perkins H.D., Howells A.J.;
 RT "The structure, sequence and developmental pattern of expression of
 RT the white gene in the blowfly *Lucilia cuprina*."
 RL Insect Mol. Biol. 5:251-260(1996).
 RN [2]
 RP SEQUENCE OF 490-584 FROM N.A.
 RX MEDLINE=90264941; PubMed=1971656;
 RA Elizur A., Vacek A.T., Howells A.J.;
 RT "Cloning and characterization of the white and topaz eye color genes
 RT from the sheep blowfly *Lucilia cuprina*."
 RL J. Mol. Evol. 30:347-358(1990).
 CC -!- FUNCTION: May be part of a membrane-spanning permease system
 CC necessary for the transport of pigment precursors into pigment
 CC cells responsible for eye color.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U38899; AAA82057.1; -.
 DR EMBL; X53265; CAA37365.1; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR005284; Pigment_permease.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR TIGRFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Pigment; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 119 126 ATP (POTENTIAL).
 FT TRANSMEM 431 451 POTENTIAL.
 FT TRANSMEM 456 476 POTENTIAL.
 FT TRANSMEM 506 526 POTENTIAL.
 FT TRANSMEM 534 554 POTENTIAL.
 FT TRANSMEM 563 583 POTENTIAL.
 FT TRANSMEM 647 667 POTENTIAL.
 SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97EED51D CRC64;

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
 RC TISSUE=Retina;
 RX MEDLINE=96256850; PubMed=8659545;
 RA Chen H.M., Rossier C., Lalioti M.D., Lynn A., Chakravarti A.,
 RA Perrin G., Antonarakis S.E.;
 RT "Cloning of the cDNA for a human homologue of the Drosophila white
 RT gene and mapping to chromosome 21q22.3.";
 RL Am. J. Hum. Genet. 59:66-75(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20408883; PubMed=10950923;
 RA Berry A., Scott H.S., Kudoh J., Talior I., Korostishevsky M.,
 RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,
 RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
 RA Antonarakis S.E., Bonne-Tamir B.;
 RT "Refined localization of autosomal recessive nonsyndromic deafness
 RT DFNB10 locus using 34 novel microsatellite markers, genomic
 RT structure, and exclusion of six known genes in the region.";
 RL Genomics 68:22-29(2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=21192304; PubMed=11279031;
 RA Porsch-Oezcueruemez M., Langmann T., Heimerl S., Borsukova H.,
 RA Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.;
 RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
 RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
 RT expression and a modulator of cellular lipid efflux.";
 RL J. Biol. Chem. 276:12427-12433(2001).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).
 RX MEDLINE=21092576; PubMed=11162488;
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
 RA Assmann G., Cullen P.;
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).

RN [6]
 RP SEQUENCE OF 33-678 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=97186700; PubMed=9034316;
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
 RA Goldenson D., Arciniegas S., Son D., Wu R.;
 RT "Isolation and characterization of a mammalian homolog of the
 RT Drosophila white gene.";
 RL Gene 185:77-85(1997).
 RN [7]
 RP INDUCTION, AND PROBABLE FUNCTION.
 RX MEDLINE=20261604; PubMed=10799558;
 RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
 RA Mangelsdorf D.J., Edwards P.A.;
 RT "Human white/murine ABC8 mRNA levels are highly induced in
 RT lipid-loaded macrophages. A transcriptional role for specific
 RT oxysterols.";
 RL J. Biol. Chem. 275:14700-14707(2000).
 RN [8]
 RP INDUCTION, AND PROBABLE FUNCTION.
 RX MEDLINE=20105556; PubMed=10639163;
 RA Klucken J., Buechler C., Orso E., Kaminski W.E.,
 RA Porsch-Oezcueruemez M., Liebisch G., Kapinsky M., Diederich W.,
 RA Drobnik W., Dean M., Allikmets R., Schmitz G.;
 RT "ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a
 RT regulator of macrophage cholesterol and phospholipid transport.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).
 RN [9]
 RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
 CC an active component of the macrophage lipid export complex. Could
 CC also be involved in intracellular lipid transport processes. The
 CC role in cellular lipid homeostasis may not be limited to
 CC macrophages.
 CC -!- SUBUNIT: May form heterodimers with several heterologous partners
 CC of the ABCG subfamily.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Predominantly
 CC localized in the intracellular compartments mainly associated with
 CC the endoplasmic reticulum (ER) and Golgi membranes.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=7;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=P45844-1; Sequence=Displayed;
 CC Name=2; Synonyms=J;
 CC IsoId=P45844-2; Sequence=VSP_000047, VSP_000051;
 CC Name=3; Synonyms=ABDE;
 CC IsoId=P45844-3; Sequence=VSP_000048, VSP_000051;
 CC Name=4; Synonyms=G;
 CC IsoId=P45844-4; Sequence=VSP_000051;
 CC Name=5; Synonyms=F;
 CC IsoId=P45844-5; Sequence=VSP_000049, VSP_000051;
 CC Name=6; Synonyms=HI;

```

CC      IsoId=P45844-6; Sequence=VSP_000046, VSP_000051;
CC      Name=7; Synonyms=C;
CC      IsoId=P45844-7; Sequence=VSP_000050, VSP_000051;
CC      -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
CC      -!- INDUCTION: Strongly induced in monocyte-derived macrophages during
CC            cholesterol influx. Conversely, mRNA and protein expression are
CC            suppressed by lipid efflux. Induction is mediated by the liver X
CC            receptor/retinoide X receptor (LXR/RXR) pathway.
CC      -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC            subfamily.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X91249; CAA62631.1; ALT_INIT.
DR      EMBL; AP001746; BAA95530.1; ALT_INIT.
DR      EMBL; AB038161; BAB13728.2; ALT_INIT.
DR      EMBL; AJ289137; CAC00730.1; ALT_INIT.
DR      EMBL; AJ289138; CAC00730.1; JOINED.
DR      EMBL; AJ289139; CAC00730.1; JOINED.
DR      EMBL; AJ289140; CAC00730.1; JOINED.
DR      EMBL; AJ289141; CAC00730.1; JOINED.
DR      EMBL; AJ289142; CAC00730.1; JOINED.
DR      EMBL; AJ289143; CAC00730.1; JOINED.
DR      EMBL; AJ289144; CAC00730.1; JOINED.
DR      EMBL; AJ289145; CAC00730.1; JOINED.
DR      EMBL; AJ289146; CAC00730.1; JOINED.
DR      EMBL; AJ289147; CAC00730.1; JOINED.
DR      EMBL; AJ289148; CAC00730.1; JOINED.
DR      EMBL; AJ289149; CAC00730.1; JOINED.
DR      EMBL; AJ289150; CAC00730.1; JOINED.
DR      EMBL; AJ289151; CAC00730.1; JOINED.
DR      EMBL; AF323658; AAK28836.1; -.
DR      EMBL; AF323644; AAK28836.1; JOINED.
DR      EMBL; AF323645; AAK28836.1; JOINED.
DR      EMBL; AF323646; AAK28836.1; JOINED.
DR      EMBL; AF323647; AAK28836.1; JOINED.
DR      EMBL; AF323648; AAK28836.1; JOINED.
DR      EMBL; AF323649; AAK28836.1; JOINED.
DR      EMBL; AF323650; AAK28836.1; JOINED.
DR      EMBL; AF323651; AAK28836.1; JOINED.
DR      EMBL; AF323652; AAK28836.1; JOINED.
DR      EMBL; AF323653; AAK28836.1; JOINED.
DR      EMBL; AF323654; AAK28836.1; JOINED.
DR      EMBL; AF323655; AAK28836.1; JOINED.
DR      EMBL; AF323656; AAK28836.1; JOINED.
DR      EMBL; AF323657; AAK28836.1; JOINED.
DR      EMBL; AF323664; AAK28842.1; -.
DR      EMBL; AF323658; AAK28833.1; -.
DR      EMBL; AF323640; AAK28833.1; JOINED.
DR      EMBL; AF323645; AAK28833.1; JOINED.
DR      EMBL; AF323646; AAK28833.1; JOINED.

```


DR EMBL; AF323647; AAK28833.1; JOINED.
 DR EMBL; AF323648; AAK28833.1; JOINED.
 DR EMBL; AF323649; AAK28833.1; JOINED.
 DR EMBL; AF323650; AAK28833.1; JOINED.
 DR EMBL; AF323651; AAK28833.1; JOINED.
 DR EMBL; AF323652; AAK28833.1; JOINED.
 DR EMBL; AF323653; AAK28833.1; JOINED.
 DR EMBL; AF323654; AAK28833.1; JOINED.
 DR EMBL; AF323655; AAK28833.1; JOINED.
 DR EMBL; AF323656; AAK28833.1; JOINED.
 DR EMBL; AF323657; AAK28833.1; JOINED.
 DR EMBL; AF323660; AAK28838.1; -.
 DR EMBL; AF323663; AAK28841.1; ALT_INIT.
 DR EMBL; AF323658; AAK28835.1; -.
 DR EMBL; AF323642; AAK28835.1; JOINED.
 DR EMBL; AF323645; AAK28835.1; JOINED.
 DR EMBL; AF323646; AAK28835.1; JOINED.
 DR EMBL; AF323647; AAK28835.1; JOINED.
 DR EMBL; AF323648; AAK28835.1; JOINED.
 DR EMBL; AF323649; AAK28835.1; JOINED.

Query Match 17.6%; Score 617; DB 1; Length 678;
 Best Local Similarity 25.7%; Pred. No. 5.5e-38;
 Matches 173; Conservative 130; Mismatches 266; Indels 104; Gaps 18;

Qy 33 DNSLYFT--YSGQPN----TLEVRDLNQVDLASQVPWFELAQFKMPWTSPSCQNSCEL 86
 ||:| :| | :| ||:| | : ||: : :
 Db 57 DNNLTEAQRFSSSLPRRAAVNIEFRDLSYSV---PEGPWWRKKGYKTL----- 100

Qy 87 GIQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWIWQPSQPQLVRK 146
 :: :| | ||::||:| || |::|::: | | : ||| | : ||
 Db 101 -LKGISGKFNSGELVAIMGPSGAGKSTLMNILAGYRETG--MKGAVLINGLPRDLRCFRK 157

Qy 147 CVAHVRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVG 206
 :: | : |||:||:| : |::| | : |::: | | ||:| |
 Db 158 VSCYIMQDDMLLPHLTVQEAMMVS AHLKLQE--KDEGRREMVKEILTALGLLSCANTRTG 215

Qy 207 NMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTL SRLAKGNRLVL 266
 : |||:|:|:| :|:| | : : ||||| | : : | : ||:| | : :
 Db 216 S-----LSGGQRKRLAIALELVNPPVMFFDEPTSGLDASCFQVVSIMKGLAQGGRSII 270

Qy 267 ISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTS 326
 ::||| : :| ||| : ::| : | | : :| | : | || | |||| : :: |
 Db 271 CTIHQPSAKLFELFDQLYVLSQGQCVRGKVCNLVPYLRDLGLNCPTYHNPADFVMEVAS 330

Qy 327 IDRRSREQEL--ATREKAQSLAALFLEKVRDL-----DDFLWKAET-----KDLD 369
 : : | | || : ||| : ||| : | |
 Db 331 GEYGDQNSRLVRVRE-----GMCDSDHKRDLGGDAEVNPF LWHRPSEEVKQTKRLKGLR 385

Qy 370 EDTCRESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIHGAEACLMSM 429
 :|: :|| || | :| : || : : :
 Db 386 KDSSSMEGCHSFSASCL-----TQFCILFKRTFLSIMRDSVLTHLRITSHIGIGL 435

Qy 430 TIGFLYFGHGSIQLSFMDTAALLF-----MIGALIPFNVILDVISKYSERAMLYYELE 483
 || || | | : : | | | ||:| : | : ||
 Db 436 LIGLLYLGIGNEAKKVLNSGFLFFSMLFLMFAALMP-----TVLTFPLE 480

Qy	484	DGL-----YTTGPGYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHLHFL	534
		: : : : :: : : : : :	
Db	481	MGVFLREHLNYWYSLKAYYLAKTMADVPFQIMFPVAYCSIVYWMTSQPSDAVRVFLFAAL	540
Qy	535	VVLVVFCCRIMALAAAAALLPTFHMAFFSNALYNSFYLAGGFMINLSSLWTPAWISKVS	594
		: : : : : : : : : : :	
Db	541	GTMTSLVAQSLGLLIGAASTSLQVATFVGPVTAIPVLLFSGFFVSFDTIPTYLQWMSYIS	600
Qy	595	FLRWCFEGLMKIQF--SRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSG	652
		::: : : : : : :::: : :	
Db	601	YVRYGFEGVILSIYGLDREDLHCDIDETCHFQKSEAILRELDVENAKLY-LDFIVLG---	656
Qy	653	GFMVLYYVSLRFI	665
		: ::	
Db	657	----IFFISLRLI	665

DR SMART; SM00382; AAA; 1.
 DR TIGREFAMs; TIGR00955; 3a01204; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Hypothetical protein; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 27 34 ATP (POTENTIAL).
 FT TRANSMEM 336 356 POTENTIAL.
 FT TRANSMEM 425 445 POTENTIAL.
 FT TRANSMEM 453 473 POTENTIAL.
 FT TRANSMEM 478 498 POTENTIAL.
 SQ SEQUENCE 598 AA; 66906 MW; 9D6414E06898E343 CRC64;

Query Match 17.1%; Score 600; DB 1; Length 598;
 Best Local Similarity 27.9%; Pred. No. 8.6e-37;
 Matches 170; Conservative 116; Mismatches 260; Indels 64; Gaps 15;

Qy 88 IQNLSFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINGQPSSPQLVRKC 147
 : |:| ||::|||:|||| |: |:|:|:| | | |:|: :: |:|
 Db 10 LHNVSGBMAESGKLLAILGSSGAGKTTLMNVLTNRNLTNLDVQGSILIDGRRANKWKIREM 69

Qy 148 VAHVRQHNLQLLPNLTVRETAFIAQMRL-PRTFSAQRDKRVEDVIAELRLRQCADTRVG 206
 | |:|: : |:| | |:|:|: : |:| |:| |:| |:| |:| |:| |:|
 Db 70 SAFVQQHDMFVGTMTAREHLQFMARLRMGDQYYSDHERQLRVEQVLTQMGLKKCADTVIG 129

Qy 207 -NMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKTLSRLAKGNRLV 265
 ::||| |:|:|:| :|:| | | | | | | | | | | | | | | | | | |
 Db 130 IPNQLKGLSCGEKKRLSFASEILTCPKILFCDEPTSGLDAFMAGHVQALRSLADNGMTV 189

Qy 266 LISLHQPRSDIFRLFDLVLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLT 325
 |:|:| | | |:| |:| | | | | | | | | | | | | | | | | | | |
 Db 190 IITIHQPSSHVYSLFNNVCLMACGRVIYLGPGDQAVPLFEKCGYPCPAYYNPADHLIRTL 249

Qy 326 SIDRRSREQELATREKAQSLAALFLEKVRDLDDFLWKAETKDLEDTC-----VES 376
 :: | : | : |:| || : || : | |
 Db 250 AVIDSDRATSMKT-----ISKIR--QGFL---STDLGQSVLAIGNANKLRAAS 292

Qy 377 SVTPLDTNCLPSPTKM-----PGAVQQFTTLIRRQISNDFRDLPTLLIHGAACLMSM 429
 || ||: || || | | | | | | | | | | | | | | | | | | | | | | | |
 Db 293 FVTGSDTS---EKTKTFFNQDYNASFWTQFLALFWRSWLTVIRDPNLLSVRLQLILITAF 349

Qy 430 TIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDV-----ISKCYSERAMLYYE 481
 | ::| | | : | : | | | : : : : | : | : |
 Db 350 ITGIVFF-----QTPVTPATIISINGIM-FNHIRNMNFMQLQFPNVPVITAEPIVIRE 401

Qy 482 LEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVLVVFC 541
 |:| | | | | : | | : | | | | | | | | | | | | | | | | | | |
 Db 402 NANGVYRTSAYFLAKNIAELPQYIILPILYNTIVYWMSSGLYPNFWNYCFASLVITILITNV 461

Qy 542 CRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTPAWISKVSFLRWCFE 601
 :: | | : : | | | | | | | | | | | | | | | | | | | | | | | |
 Db 462 AISISYAVATIFANTDVAMTILPIFVVPIMAFGGFFITFDAIPSYFKWLSSLSYFKYGYE 521

Qy 602 GLM-----KIQFSRRTYKMPLGNLTI---AVSGDKILSAMELD-SYPLYAIYLIVIGLSG 652
 | : : : : : : | : | : | : | : | : | : | : | : | : | : |
 Db 522 ALAINEWDSIKVIPECFNSSMTAFALDSCPKNGHQVLESIDFSASHKIFDI-SILFGMFI 580

Qy 653 GFMVLYYVSL 662

| :: ||:|
Db 581 GIRIIAYVAL 590

Search completed: February 27, 2004, 07:12:40
Job time : 11.4203 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:43 ; Search time 37.3606 Seconds
(without alignments)
5683.620 Million cell updates/sec

Title: US-09-989-981A-8
Perfect score: 3506
Sequence: 1 MAGKAAEERGLPKGATPQDT.....FMVLYYVSLRFIKQKPSQDW 673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query		Description	
No.	Score	Match	Length DB ID		

1	2883.5	82.2	672	11	Q7TSR6	Q7tsr6	mus	musculu
2	2877.5	82.1	672	11	Q7TSR7	Q7tsr7	mus	musculu
3	2871	81.9	673	11	Q8R543	Q8r543	mus	musculu
4	2835.5	80.9	672	11	Q8CIQ5	Q8ciq5	rattus	norv
5	742	21.2	668	10	Q9ARU4	Q9aru4	oryza	sativ
6	739.5	21.1	672	10	Q9LI82	Q9li82	arabidopsis	
7	735.5	21.0	725	10	Q9ZU35	Q9zu35	arabidopsis	
8	735.5	21.0	725	10	Q9ASR9	Q9asr9	arabidopsis	
9	730.5	20.8	648	10	Q9C6W5	Q9c6w5	arabidopsis	
10	723.5	20.6	646	10	Q9C6R7	Q9c6r7	arabidopsis	
11	709	20.2	662	10	Q949Y4	Q949y4	arabidopsis	
12	708	20.2	662	10	Q84TH5	Q84th5	arabidopsis	
13	700	20.0	609	10	Q9C8W6	Q9c8w6	arabidopsis	
14	695.5	19.8	801	5	Q8T691	Q8t691	dictyosteli	
15	686.5	19.6	652	11	Q7TSR8	Q7tsr8	mus	musculu
16	668.5	19.1	737	10	Q9FT51	Q9ft51	arabidopsis	
17	667	19.0	657	11	Q7TMS5	Q7tms5	mus	musculu
18	666	19.0	657	11	Q9R004	Q9r004	mus	musculu
19	665	19.0	687	5	Q9NH94	Q9nh94	bombyx	mori
20	659	18.8	751	10	Q93YS4	Q93ys4	arabidopsis	
21	658	18.8	687	5	Q94960	Q94960	drosophila	
22	657	18.7	657	11	Q80W57	Q80w57	rattus	norv
23	657	18.7	657	11	Q80ST1	Q80st1	rattus	norv
24	653	18.6	656	6	Q8MIB3	Q8mib3	sus	scrofa
25	653	18.6	657	11	Q80XF3	Q80xf3	rattus	norv
26	651.5	18.6	635	10	Q9SZR9	Q9sZR9	arabidopsis	
27	647.5	18.5	679	5	Q8IS30	Q8is30	bactrocera	
28	643.5	18.4	643	5	Q7YYX5	Q7yyx5	cryptospori	
29	643	18.3	670	5	O77423	O77423	bactrocera	
30	642.5	18.3	655	4	Q8IX16	Q8ix16	homo	sapien
31	642.5	18.3	655	4	Q96TA8	Q96ta8	homo	sapien
32	642.5	18.3	679	5	Q9BH97	Q9bh97	ceratitidis	c
33	634.5	18.1	655	4	Q96LD6	Q96ld6	homo	sapien
34	621.5	17.7	567	10	Q9FG17	Q9fg17	arabidopsis	
35	621	17.7	666	11	Q9EPG9	Q9epg9	rattus	norv
36	620	17.7	662	4	Q86SU8	Q86su8	homo	sapien
37	617.5	17.6	669	5	Q8WRF2	Q8wrf2	tribolium	c
38	617	17.6	692	10	Q7XUM2	Q7xum2	oryza	sativ
39	616.5	17.6	691	10	Q8RWI9	Q8rwi9	arabidopsis	
40	613	17.5	594	10	Q9LJC3	Q9ljc3	arabidopsis	
41	612.5	17.5	669	5	Q8WRR1	Q8wrr1	tribolium	c
42	610.5	17.4	609	5	Q9VQN4	Q9vqn4	drosophila	
43	607.5	17.3	703	10	Q8RXN0	Q8rxn0	arabidopsis	
44	598.5	17.1	785	4	Q96L76	Q96l76	homo	sapien
45	597.5	17.0	692	5	P91892	P91892	aedes	aegyp

ALIGNMENTS

RESULT 1

Q7TSR6

ID Q7TSR6 PRELIMINARY; PRT; 672 AA.
AC Q7TSR6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

DE      ATP-binding cassette sub-family G member 8.
GN      ABCG8.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Pera/Ei; TISSUE=Liver;
RA      Wittenburg H., Lyons M.A., Li R., Churchill G.A., Carey M.C.,
RA      Paigen B.;
RT      "Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone
RT      Susceptibility: Evidence from a Cross of Pera/Ei and I/Ln Inbred
RT      Mice.";
RL      Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AY196216; AAO45096.1; -.
KW      ATP-binding.
SQ      SEQUENCE 672 AA; 75867 MW; CAB720502EA8FE21 CRC64;

```

Qy	1	MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSYGQPN TLEVRDLNYQVDLAS	60
Db	1	MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYFTYSYGQSN TLEVRDLTYQVDIAS	60
Qy	61	QVPWFQEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQGMLAIIGSSGCGRASLLDVITG	120
Db	61	QVPWFQEQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSQGMLAIIGSSGCGRASLLDVITG	120
Qy	121	RGHGGKIKSGQIWINGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFS	180
Db	121	RGHGGKMKSGQIWINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRETAFIAQMRLPRTFS	180
Qy	181	QAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNP GILILDEPT	240
Db	181	QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPT	240
Qy	241	SGLDSFTAHLNVLKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHM	300
Db	241	SGLDSFTAHLNVLTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQOM	300
Qy	301	VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL	360
Db	301	VQYFTSIGHPCPRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFL	360
Qy	361	WKAETKDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH	420
Db	361	WKA EAKELNTSTHTVSLTLTQD TDC-GTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIH	419
Qy	421	GAEACLMSMTIGFLYFGHSGIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYY	480
Db	420	GSEACLMSLIIGFLYFGHGAQQLSFMDTAALLFMIGALIPFNVILDVVSKCHSERSMLYY	479
Qy	481	ELEDGLYTTGPYFFAKILGELPEHCAYII IYGMPTYWLANLRPLGQPFLHLHFLVLVVF	540

Db 480 ELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWL VVF 539

QY 541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF 600
 ||| |||||:|:|||||:||| ||||| ||||| :|| |||||:|||||

Db 540 CCRTMALAASAMLPTFHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWCF 599

QY 601 EGLMKIQFSRRTYKMPGLNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV 660
 |||:||||: | :|| | :: || :|||:|:|:||||| |||: |||:

Db 600 SGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLLYYL 659

QY 661 SLRFIKQKPSQDW 673
 ||: |||| |||

Db 660 SLKLIKQKSIQDW 672

RESULT 2

Q7TSR7

ID Q7TSR7 PRELIMINARY; PRT; 672 AA.

AC Q7TSR7;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ATP-binding cassette sub-family G member 8.

GN ABCG8.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=I/LnJ; TISSUE=Liver;

RA Wittenburg H., Lyons M.A., Li R., Churchill G.A., Carey M.C.,

RA Paigen B.;

RT "Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone

RT Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred

RT Mice.";

RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AY196215; AA045095.1; -.

KW ATP-binding.

SQ SEQUENCE 672 AA; 75805 MW; E5B30B5890200A41 CRC64;

Query Match 82.1%; Score 2877.5; DB 11; Length 672;

Best Local Similarity 81.7%; Pred. No. 3.6e-215;

Matches 550; Conservative 52; Mismatches 70; Indels 1; Gaps 1;

QY 1 MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLAS 60
 || | || | | || ||||| ||||| ||||| ||||| |||||:||

Db 1 MAEKTKEETQLWNGTVLQDASGLQDSLFSSES DNSLYFTYSGQSNTLEVRDLTYQVDIAS 60

QY 61 QVPWFQQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITG 120
 ||||| |||||:|| | | |:|||||:|||||:|||||:|||||:|||||:|||||

Db 61 QVPWFQQLAQFKIPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITG 120

QY 121 RGHGGKIKSGQIWINGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFS 180
 |||||:||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||

Db 121 RGHGGKMKSGQIWINGQPSTPQLVRKCVAHVRQHDQLLPNLTVRETAFIAQMRLPRTFS 180

Db 420 HGSEACLSLIIGFLYYGHGAKQLSFMDTAALLFMIGALIPFNVILDVVSCHSERSMLY 479

Qy 480 YELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVWLVV 539
 ||||| : |||||

Db 480 YELEDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELFLHFLLVWLVV 539

Qy 540 FCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWC 599
 |||| : |||||

Db 540 FCCRNMALAASAMLPTFHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRWC 599

Qy 600 FEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYY 659
 | ||| : ||| : || : || : ||| : ||| : ||| : |||

Db 600 FSGLMQIQFNHLYTTQIGNFTFSILGDTMISAMDLSHPLYAIYLIVIGISYGFLLYY 659

Qy 660 VSLRFIKQKPSQDW 673
 : || : |||| |||

Db 660 LSLKLIKQKSIQDW 673

RESULT 4

Q8CIQ5

ID Q8CIQ5 PRELIMINARY; PRT; 672 AA.

AC Q8CIQ5;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Sterolin 2.

GN ABCG8.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley;

RA Yu H., Lu K., Lee M., Pandit B., Patel s.B.;

RT "The rat Abcg5 and Abcg8: characterization, chromosomal assignment and

RT genetic variation in sitosterolemic rats.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AY145899; AAN64276.1; -.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR003439; ABC_transporter.

DR Pfam; PF00005; ABC_tran; 1.

DR ProDom; PD000006; ABC_transporter; 1.

DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.

DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.

SQ SEQUENCE 672 AA; 75906 MW; 2FE0846E71BD9D47 CRC64;

Query Match 80.9%; Score 2835.5; DB 11; Length 672;

Best Local Similarity 79.9%; Pred. No. 6.7e-212;

Matches 538; Conservative 57; Mismatches 77; Indels 1; Gaps 1;

Qy 1 MAGKAAEERGLPKGATPQDTSGLQDRLFSSES DNSLYFTYSQGPNTLEVRDLNYQVDLAS 60
 || | || | || | ||| : ||||| ||||| |||||

Db 1 MAEKTKEETQLWNGTVLQDASSLQDSVFSSSEDNSLYFTYSGQSNLTLEVRDLTYQVDMAS 60

Qy 61 QVPWFEEQLAQFKMPWTSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITG 120
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 61 QVPWFEEQLAQFKLPWRSRGSQDSWDLGIRNLSFKVRSQMLAIIGSAGCGRATLLDVITG 120

Qy 121 RGHGGKIKSGQIWINQPPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFS 180
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 121 RDHGGKMKSGQIWINQPPSTPQLIQKCVAHVRQQDQLLPNLTVRETFTFIAQMRLPKTFS 180

Qy 181 QAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPT 240
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 181 QAQRDKRVEDVIAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPT 240

Qy 241 SGLDSFTAHLNVLKTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAQAQM 300
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 241 SGLDSFTAHLNVLRTLRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAQAQM 300

Qy 301 VQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFL 360
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 301 VQYFTSIGYPCPRYSNPADFYVDLTSIDRRSKEQEVATMEKARLLAALFLEKVRQGFDDFL 360

Qy 361 WKAETKDLDEDTCESSVTPLDTNCLPSPTKMPGAVQQFTTLIRRQISNDFRDLPTLLIH 420
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 361 WKAEAKSLDTGTAVSQTLTQDTNC-GTAAELPGMIQQFTTLIRRQISNDFRDLPTLFIH 419

Qy 421 GAEACLMSMTIGFLYFGHSGSIQLSFMDTAALLFMIGALIPFNVILDVISKYSERAMLYY 480
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 420 GAEACLMSLIIGFLYFGHADKPLSFMDMAALLFMIGALIPFNVILDVSKCHSERSLLYY 479

Qy 481 ELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLLHFLLVWLWVF 540
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 480 ELEDGLYTAGPYFFAKVLGELPEHCAYVIIYGMPIYWLTLNLRPGPELFLHFMLLWLVVF 539

Qy 541 CCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCF 600
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 540 CCRTMALAASAMLPFTHMSSFCCNALYNSFYLTAGFMINLNNLWIVPAWISKMSFLRWCF 599

Qy 601 EGLMKIQFSRRTYKMPGLNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV 660
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 600 SGLMQIQFNHGYTTQIGNLTFVPGDAMVTAMDLSHPLYAIYLIVIGISCGFLSLYYL 659

Qy 661 SLRFIKQKPSQDW 673
| | : | | | | | | |

Db 660 SLKFIKQKSIQDW 672

RESULT 5

Q9ARU4

ID Q9ARU4 PRELIMINARY; PRT; 668 AA.
AC Q9ARU4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC transporter.
GN P0445D12.3.
OS Oryza sativa (Rice).

```

OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC      Ehrhartoideae; Oryzeae; Oryza.
OX      NCBI_TaxID=4530;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Nipponbare;
RA      Sasaki T., Matsumoto T., Yamamoto K.;
RT      "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT      clone:P0445D12.";
RL      Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC      -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR      EMBL; AP003046; BAB40032.1; -.
DR      Gramene; Q9ARU4; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR      GO; GO:0000166; F:nucleotide binding; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW      ATP-binding; Transport.
SQ      SEQUENCE 668 AA; 73368 MW; D1875B8C75B0F3B2 CRC64;

```

Qy	75	WTSPSCQNSCELG-----IQNLSFKVRSQMLAIIGSSGCGRASLLDVITGR--GHGGK	126
Db	58	WARITCALKNKRGDVARFLLSNASGEAKSGRLLALMGPSGSGKTTLLNLVLAGQLTASPSL	117
Qy	127	IKSGQIWINQQPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDK	186
Db	118	HLSGFLYINGRPISEGQYK--IAYVRQEDLFFSQLTVRETLSLAELQLRRTLTPERKES	175
Qy	187	RVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSE	246
Db	176	YVNDLLFRLGLVNCADSIVGDAKVRGISGGEKKRLSLACELIASPSIIFADEPTTGLDAF	235
Qy	247	TAHNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLG-AAQHMVQYFT	305
Db	236	QAEKVMETLRLQLAEDGHTVICSIHQPRGSVYGKFDIVLLSEGEVIYMGPAKEEPLLYFA	295
Qy	306	AIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFLWKAET	365
Db	296	SLGYHCPDHVNPAEFLADLISVDYSSAESVQSSRKRIENLIEEFSNKV-----AIT	346
Qy	366	KDLDEDTCESSVT-PLDTNCLP-----SPTK-MPGAVQQFTTLIRRQISNDFRDLPTLL	418
Db	347	ES-----NSSLTNPEGSEFSPKLIQKSTTKHRRGWWRQFRLLFKRAWMQAFRDGPTNK	399

Qy 419 IHGAEACLSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAML 478
 : : :: | :: | | | | : : : ||::
 Db 400 VRARMSVASAIIFGSVFWRMGKTQTSIQDRMGLLQVTAINTAMAALTKTVGVFPKERAIV 459
 Qy 479 YYELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLLVWL 538
 | | | || :|:| |:| |:| :|:| | :| | :| :
 Db 460 DRERAKGSYALGPYLSSKLLAEIPIGAAPFLIFGSILYPMSKHLPTFSRFAKFCGIVTVE 519
 Qy 539 VFCCRIMALAAAALLPTFHMAFFSNALYNSFYLAGGFMINLSSLWTVPAWISKVSFLRW 598
 | | | |:| | :| | :|:| : : | | || :|
 Db 520 SFAASAMGLTVGAMAPTTEAAMALGPSLMTVFIVFGGYVNPDPNTPVIFRWIPKVSIRW 579
 Qy 599 CFEGLMKIQF-----SRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSG 652
 |:| :| : :| : | : :| | : :
 Db 580 AFQGLCINEFKGLQFEQQHSYDIQTGE-----QALERFSLGGIRIADTLVAQ 626
 Qy 653 GFMVLYYVSLRFI---KQKP 669
 | :::: | : | :|
 Db 627 GRILMFYWLTYLLLKKNRP 646

RESULT 6

Q9LI82

ID Q9LI82 PRELIMINARY; PRT; 672 AA.
 AC Q9LI82;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ABC transporter-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL; AP001313; BAB03081.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR InterPro; IPR006162; Ppantne_S.

DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
 KW ATP-binding.
 SQ SEQUENCE 672 AA; 75269 MW; 20B2D99215600135 CRC64;

Query Match 21.1%; Score 739.5; DB 10; Length 672;
 Best Local Similarity 30.6%; Pred. No. 7.9e-49;
 Matches 221; Conservative 126; Mismatches 251; Indels 125; Gaps 27;

Qy	7	EERGLPK--GATPQDTSGLQDRLFSSES-----DN-----SLYFTYSGQPN	45
		: : : : :	
Db	7	QESSFPKTPSANRHETSPVQENRFSSPSHVNPCLDDNDHDGSPHQSRQSSVLRQSLRPI	66
Qy	46	TLEVRDLNYQVD-----LASQVPWFEEQLAQFKMPWTSPSCQNSCELGIQNLSFKVR	96
		: : : : :	
Db	67	ILKFEELTYSIKSQTGKGSYWFGSQEPKPNRLVL-----KCVSGI-----VK	108
Qy	97	SGQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSSPQLVRKCAVHRQHNQ	156
		: : : : : : : : : : : :	
Db	109	PGELLAMLGPSSGSKTTLVTALAGRLQ-GKL-SGTVSYNGEPTSSVKRK-TGFVTQDDV	165
Qy	157	LLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGG	216
		: : : : :: :: : : : : :: : : :	
Db	166	LYPHLTMETLTYTALLRLPKELTRKEKLEQVEMVSDLGLTRCCNSVIGGGILIRGISGG	225
Qy	217	ERRRVSIGVQLLWNPgililDEPTSGLDsftahnlVKtLSRLAKGNRLVLISLHQPRSDI	276
		: :: : : : : : : :	
Db	226	ERKRVSIGQEMLVNPSLLLLDEPTSGLDSTTAARIVATLRSLARGGRTVVTTHIQPSSRL	285
Qy	277	FRLFDLVLLMTSGTPIYLGAQHMVQYFTAIGY-PCPRYSNPADFYVDLT---SIDRRSR	332
		: : :: : :: : : : : :	
Db	286	YRMFDKVLVLSEGCPIYSGDSGRVMEYFGSIGYQPGSSFVNPADFVLDLANGITSDTKQY	345
Qy	333	EQELATREKAQSLAALFLEKVRDLDDFLWKAETKDLDDETCVSSVT-PLD-----	382
		: : : : : : :	
Db	346	DQ-IETNGRLDR-----LEEQNSVKQSLISSYKKNLYPPLKEEVSRTPQDQTNARLRKK	399
Qy	383	--TNCLPSPTKMPGAVQOFTTLIRRQIS---NDFRDLPTLLIHGAELMSMTIGFLYF	436
		: : : : : : : :	
Db	400	AITNRWPTSWM-----QFSVLLKRGKERSHESFGLRIFMVMS-----VSLLSGLLWW	449
Qy	437	GHGSIQLSEMDTAALLFMIGALIPFNVILDVISKCYSERAMLYYELEDGLYTTGPYFFAK	496
		: : : : : :	
Db	450	-HSRV-AHLQDQVGLLFFFSIFWGFFPLFNAIETFPQERPMLIKERSSGIYRLSSYYIAR	507
Qy	497	ILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLLHFLLVLVVFCRIMALAAAALLPTF	556
		: : : : : : : : : :	
Db	508	TVGDLPMELILPTIFVTITYWMGGLKPSLTTFIMTLMIVLYNVLVAQGVGLALGAILMDA	567
Qy	557	HMAFFSNALYNSFYLAGGFMINLSSLWTVP---AWISKVSFLRWCFEGLMKIQFS-RRT	612
		: : : : : : : : : :	
Db	568	KKAATLSSVLMVLVFLLAGGYIQ-----HIPGFIAWLKYSFVSHYCYKLLVGVQYTWDEV	622

Qy 613 YKMPLG-----NLTIAVSGDKILSAMELDSYPLYAIYLIVIGLSGGFMVLYYV 660
 | : | | | : : : | | : : : | : || | :
 Db 623 YECGSLHCSVMDYEGIKNLRI---GNMMWDVLAL-----AVMLLL-----YRVLAYL 667

Qy 661 SLR 663
 : ||
 Db 668 ALR 670

RESULT 7

Q9ZU35

ID Q9ZU35 PRELIMINARY; PRT; 725 AA.
 AC Q9ZU35;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative ABC transporter.
 GN AT2G01320.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana."
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AC006200; AAD14532.1; -.
 DR PIR; C84423; C84423.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.

KW ATP-binding; Transport.

SQ SEQUENCE 725 AA; 78899 MW; 7DB2E556FE3553D7 CRC64;

Query Match 21.0%; Score 735.5; DB 10; Length 725;

Best Local Similarity 30.0%; Pred. No. 1.8e-48;

Matches 186; Conservative 123; Mismatches 229; Indels 81; Gaps 15;

```
Qy      75 WTSPSC-----QNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITG-----RG 122
      | : :|      |      ::||:| : : |::|||:| || |: :||:|: |      |
Db      72 WRNITCSLSDKSSKSVRFLKKNVSGEAKPGRLLAIMGPSGSGKTTLLNVLAGQLSLSPRL 131

Qy     123 HGGKIKSGQIWINGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQA 182
      |      || : :||:||| : : : | ||| :      |||||:| |:::| |      |
Db     132 H-----LSGLLEVNGKPSSSKAYK--LAFVRQEDLFFSQLTVRETLSFAAELQLPEISSAE 185

Qy     183 QRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSG 242
      :||: | ::: :| | |||: ||: |||:||||:|:|: :|: :| :: |||:|
Db     186 ERDEYVNNLLLKLGLVSCADSCVGDAKVRGISGGEKKRLSLACELIASPSVIFADEPTTG 245

Qy     243 LDSFTAHNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLG-AAQHMV 301
      ||:| | :::| | ||: | : | :||| : : || :|:| | | : | :
Db     246 LDAFQAEKVMETLQKLAQDGH TVICS IHQPRGSVYAKFDDIVLLTEGTLVYAGPAGKEPL 305

Qy     302 QYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFLW 361
      || | : || : |||:| || |:| | | :::: :| | : :
Db     306 TYFGNFGFLCPEHVNPAEFLADLISVDYSSSETVYSSQKRVHALVD AFSQR----- 356

Qy     362 KAETKDLDEDTCEVSSV---TPLDTNCLPSPTK-----MPGAVQQFTTLIRR 405
      ||| ||| : ||      | :|| |::|
Db     357 -----SSSVLYATPLS---MKEETKNGMRPRRKAIVERTDGWWRQFFLLLR 400

Qy     406 QISNDFRDLPTLLIHGAACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVIL 465
      || || : : : :| ::: | | | | || :|: | :
Db     401 AWMQASRDGPTNKVRARMSVASAVIFGSVFWRMGKSQTSIQDRMGLL-QVAAI---NTAM 456

Qy     466 DVISKCY----SERAMLYYELEDGLYTTGPFYFFAKILGELPEHCAYIIIIYGMPTYWLANL 521
      :| | |||: | | |: ||| :| : |:| | : :| | :| |
Db     457 AALTKTVGVFPKERAI VDRERSKGSYS LGPYLLSKTIAEIPIGAAPPLMFGAVLYPMARL 516

Qy     522 RPGLQPFLHLHLLVVLVVFCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLS 581
      | | | :| : | | | |::|: | : | : ||: :|
Db     517 NPTLSRFGKFCGIVTVESFAASAMGLTVGAMVPSTEAMAVGPSLMTVFIVFGGYVNAD 576

Qy     582 SLWTVPAWISKVSFLRWCFEGLMKIQFS-----RRTYKMPLGNLT---IAVSGDKILSA 632
      : : || : | :|| |:|| :|| :| : : | : : | :|
Db     577 NTPIIFRWIPRASLIRWAFQGLCINEFSGLKFDHQNTFDVQTGEQALERLSFGGRRIRET 636

Qy     633 MELDSYPLY----AIYLIV 647
      : | | | ||:
Db     637 IAAQSRILMFWYSATYLLL 655
```

RESULT 8

Q9ASR9

ID Q9ASR9 PRELIMINARY; PRT; 725 AA.

AC Q9ASR9;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```

DT      01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      At2g01320/F10A8.20.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA      Bowser L., Carninci P., Chung M.K., Goldsmith A.D., Hayashizaki Y.,
RA      Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA      Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M.,
RA      Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T., Satou M.,
RA      Seki M., Southwick A., Toriumi M., Yamada K., Yu G., Shinozaki K.,
RA      Davis R.W., Theologis A., Ecker J.R.;
RT      "Arabidopsis cDNA clones.";
RL      Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Cheuk R., Chen H., Kim C.J., Shinn P., Banh J., Bowser L.,
RA      Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA      Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA      Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA      Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA      Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA      Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT      "Arabidopsis ORF clones.";
RL      Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
CC      -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR      EMBL; AF367318; AAK32905.1; -.
DR      EMBL; AY133617; AAM91447.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW      ATP-binding; Transport.
SQ      SEQUENCE 725 AA; 78998 MW; 68A7E556FE2FE3D7 CRC64;

```

QY 183 QRDKRVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILILDEPTSG 242
 :||: | :: :| | |||: ||: |||:||||:|:|: :|: :| :: |||:|
 Db 186 ERDEYVNNLLKLGLVSCADSCVGDVGRGISGGEKKRLSLACELIASPSVIFADEPTTG 245

QY 243 LDSFTAHNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLG-AAQHMV 301
 ||:| | ::|| :||: | :|:|||| :: || :|:| || :| | | : :
 Db 246 LDAFQAEKVMETLQKLAQDGHGTVICSIHQPRGSVYAKFDDIVLLTEGTLVYAGPAGKEPL 305

QY 302 QYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKAQSLAALFLEKVRDLDDFLW 361
 || | : || : |||:| || |:| | | :|:| :| | :|
 Db 306 TYFGNFGFLCPEHVNPAEFLADLISVDYSSSETVYSSQKRVHALVDAFSQR----- 356

QY 362 KAETKDLDEDTCEVSSV---TPLDTNCLPSPTK-----MPGAVQQFTTLIRR 405
 ||| ||| : || | :|| |::|
 Db 357 -----SSSVLYATPLS---MKEETKNGMRPRRKAIVERTDGWWRQFFLLLR 400

QY 406 QISNDFRDLPTLLIHGAACLMSMTIGFLYFGHGSIQLSFMDTAALLFMIGALIPFNVL 465
 || || : : :: | ::| | | | || :|: | :
 Db 401 AWMQASRDGPTNKVRARMSVASAVIFGSVFWRMGKSQTSIQDRMGLL-QVAAI---NTAM 456

QY 466 DVISKCY----SERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANL 521
 ::| |||:| | |:| || :| :|:| |: ::| | :| |
 Db 457 AALTKTVGVFPKERAIVDRERSKGSYSLGPYLLSKTIAEIPIGAAPLMFGAVLYPMARL 516

QY 522 RPGLQPFLHLHLLVWLVVFFCCRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLS 581
 | | | :| : | | | :|:| | :| | :|:| :|
 Db 517 NPTLSRFGKFCGIVTVESFAASAMGLTVGAMVPSTEAMAVGPSLMTVFIVFGGYVNAD 576

QY 582 SLWTVPAWISKVSFLRWCFEGLMKIQFS-----RRTYKMPLGNLT---IAVSGDKILSA 632
 : : || :| :|| |:|| :|| :|: :| :| :| :|
 Db 577 NTPIIFRWIPRASLIRWAFQGLCINEFSGLKFDHQNTFDVQTGEQALERLSFGGRRIRET 636

QY 633 MELDSYPLY----AIYLIV 647
 : | | | ||:|
 Db 637 IAAQSRILMFWYSATYLLL 655

RESULT 9

Q9C6W5

ID Q9C6W5 PRELIMINARY; PRT; 648 AA.
 AC Q9C6W5;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (ABC transporter, putative).
 GN F27M3_2 OR AT1G31770/F27M3_2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,

RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant *Arabidopsis*
 RT *thaliana*.";
 RL Nature 408:816-820(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 RT annotation.";
 RL Genome Biol. 0:0-0(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.;
 RT "Full-Length cDNA from *Arabidopsis thaliana*.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
 RA Hayashizaki Y., Shinozaki K.;
 RT "*Arabidopsis thaliana* full-length cDNA.";
 RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC074360; AAG60152.1; -.
 DR EMBL; AY088793; AAM67104.1; -.
 DR EMBL; AK117530; BAC42192.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 648 AA; 72618 MW; D52A2D2434A5BB9D CRC64;

Query Match 20.8%; Score 730.5; DB 10; Length 648;
 Best Local Similarity 30.7%; Pred. No. 3.8e-48;
 Matches 211; Conservative 117; Mismatches 269; Indels 91; Gaps 19;

Qy 9 RGLPKGATPQDTSGLQDRLFSSES--DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFE 66
 :||| : | | | :|: | :| | | : : | : | : |
 Db 19 QGLPDMSDTQSKSVLAFPTITSQPGLQMSMY-----PITLKFEVVYKVKI-----E 65

Qy 67 QLAQFKMPWTSPSCQNSCELGIQNLSEFKVRSQMLAIIGSSGCGRASLLDVITGRGHGK 126
 | :| | | : : : | | : ||:| | | : || : | | |
 Db 66 QTSQCMGSWKSKE-----KTILNGITGMVCPGEFLAMLGPSGSGKTTLLSALGGR--LSK 118

Qy 127 IKSGQIWINQPPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDK 186
 ||:| ||| | : | : | | : | | :||| || | | :||| : : : :
 Db 119 TFSGKVMYNGQPFSGCIKRR-TGFVAQDDVLYPHLTVWETLFFTALLRPLSSSLTRDEKAE 177

Qy 187 RVEDVIAELRLRQCADTRVGNMYVRGLSGGERRRVSVIGVQLLWNPGLILDEPTSGLDSE 246
 | : |||| | :| : : | | : |||| : |||| : | || : | : |||||
 Db 178 HVDRVIAELGLNRCTNSMIGGPLFRGISGGEKRVSVIGQEMLINPSLLLLDEPTSGLDST 237

Qy 247 TAHNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTA 306
 ||| :| | : ||| | | : : ||| | | : : || | : : | : ||| ||| | : || : :
 Db 238 TAHRIVTTIKRLASGGRTVVTTHIQPSSRIYHMFQKVVLLSEGSPIIYGAASSAVEYFSS 297

Qy 307 IGYPCPRYSNPADFFYVDLTS-----IDRRSREQELATREKAQSLAALFEKVRDLDDFLW 361
 :| : |||| : || : : : ||| | : : || : : : :
 Db 298 LGFSTSLTVNPADLLLDLANGIPPDTQKETSESEQKTVK--ETLV SAYEKNI----- 347

Qy 362 KAETKDLDEDTCVESS----VTPLDTNCLPSPTKMPGAVQQFTTLIRROI-SNDFRDLPT 416
 || | : | | | | | | | | | | | | : | : | : |
 Db 348 --STK-LKAELCNAESHSYEYTKAAAKNLKSEQWCTTWWYQFTVLLQGRVRRERFESFNK 404

Qy 417 LLIHGAEACLMSMTIGFLYFGHSGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERA 476
 | | : : | | : | : : | ||| | : : : | :
 Db 405 LRIF---QVISVAFLGGLLWWH-TPKSHIQDRTALLFFFSVFWGFPYLYNAVFTFPQEK 460

Qy 477 MLYYELEDGLYTTGPYFFAKILGELPEHCAYIIYGMPTYWLANLRPGLQPFLHFLVW 536
 || | | :| || | : : || | | : || : | | : | | : |
 Db 461 MLIKERSSGMYRLSSYFMARNVGDLPLELALPTAFVFIIYWMGGLKPDPTTFILSLLVVL 520

Qy 537 LVVFCCRIMALAAAAALLPTFHMAFFSNALYNSFYLAGGFMINLSSLWTVP---AWISKV 593
 | : : || ||| | : : | : || : : : | | : :
 Db 521 YSVLVAQGLGLAFGALLMNIKQATTLASVTTLVFLIAGGYVQ-----QIPPFIVWLKYL 575

Qy 594 SFLRWCFEGLMKIQFSRRTY-----KMPLGNLTIAVSGDKILSAMEL 635
 | : : || : | : || : | | | | | | : : |
 Db 576 SYSYYCYKLLGLGIQYTDDDYECSEKGVWCRVGFPAIKSMGLNNLWI----DVFVMGVML 631

Qy 636 DSYPLYAIYLIVIGLSGGFMVLYYVSLR 663
 | | | : | | : | | |
 Db 632 VGYRLMA-----YMALHRVKLR 648

RESULT 10

Q9C6R7

ID Q9C6R7 PRELIMINARY; PRT; 646 AA.
 AC Q9C6R7;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ABC transporter, putative.

GN F5M6.22.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana."
 RL Nature 408:816-820(2000).
 DR EMBL; AC079041; AAG50724.1; -.
 DR PIR; C86441; C86441.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding.
 SQ SEQUENCE 646 AA; 72342 MW; 7A9624F82FD88A6E CRC64;

Query Match 20.6%; Score 723.5; DB 10; Length 646;
 Best Local Similarity 30.6%; Pred. No. 1.3e-47;
 Matches 208; Conservative 119; Mismatches 262; Indels 91; Gaps 20;

Qy 22 GLQDRLFSSSESDNSLYF-TYSGQPN-----TLEVRDLNYQVDLASQVPWFQEQLAQFKMP 74
 || | : ::| : | | : || : :::: |:| : || :|
 Db 20 GLPD-MSDTQSKSVLAFPTITSQPGLQMSMPITLKEVVYKVKI-----EQTSQCMGS 71
 Qy 75 WTSPSCQNSCELGIQNLSEFKVRSGQMLAIIGSSGCGRASLLDVITGRGHGKIKSGQIWI 134
 | | : : : : | | : ||:| || : :|| : || | ||::
 Db 72 WKSKE-----KTILNGITGMVCPGEFLAMLGPSGSGKTTLLSALGGR--LSKTFSGKVMY 124

Qy 135 NGQPSSPQLVRKCVAHVRQHNQLLPNLTVRETAFIAQMRLPRTFSQAQRDKRVEDVIAE 194
 ||| | : | : | | : | | : || | || | | : || | : : : : | : |||
 Db 125 NGQPFSGCIKRR-TGFVAQDDVLYPHLTWETLFFTALLRLPSSLTRDEKAEHVDRVIAE 183

Qy 195 LRLRQCADTRVGNMYVRGLSGGERRRVSGVQLLWNPGLILDEPTSGLDSTAHNLVKT 254
 | | : | : : | | : ||| : : ||| : : | | : | : ||| ||| || | : | |
 Db 184 LGLNRCTNSMIGGPLFRGISGGEKKRVSIGQEMLINPSLLLLDEPTSGLDSTTAHRIVTT 243

Qy 255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314
 : ||| | | | : : ||| | | : : | | : : : | : ||| ||| | : || : : : :
 Db 244 IKRLASGGRTVVTTIHQPSSRIYHMFDKVLLSEGSPYIYGAASSAVEYFSSLGFSTSLT 303

Qy 315 SNPADFYVDLTS-----IDRRSREQELATREKAQSLAALFLEKVRDLDDFLWKAETKDLD 369
 |||| : || : : : ||| | : : | : : : : || |
 Db 304 VNPADLLLDLANGIPPDTQKETSEQEQTVK--ETLVSAYEKNI-----STK-LK 350

Qy 370 EDTCCESS-----VTPLDTNCLPSPTKMPGAVQQFTTLIRRQI-SNDFRDLPTLLIHGAEA 424
 : | | | | | | | | | | | | | | | | : | : | : | | | |
 Db 351 AELCNAESHSEYETKAAAKNLKSEQWCTTWYQFTVLLQRGVRERRFESFNKLRIF---Q 407

Qy 425 CLMSMTIGFLYFGHSGSIQLSFMDTAALLFMIGALIPFNVILDVISKCYSERAMLYYELED 484
 : : | | : | : : | | | | | | | | : : : : | : | | |
 Db 408 VISVAFLGGLLWWH-TPKSHIQDRTALLFFFSVFWGFYPLYNVFTFPQEKRMILIKERS 466

Qy 485 GLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVWLVFCCRI 544
 | : | | | : : | : | | | : | | : | | | : | | : | :
 Db 467 GMYRLSSYFMARNVGDLPLELALPTAFVFIIYWMGGLKPDPTTFILSLLVLVLYSVLVAQG 526

Qy 545 MALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVP---AWISKVSFLRWCFE 601
 : || ||| | : : : | : || : : : | : | : : : : : :
 Db 527 LGLAFGALLMNIKQATTLASVTTLVFLIAGGYVQ-----QIPPFIVWLKYLSSYYCYK 581

Qy 602 GLMKIQFSRRTY-----KMPLGNLTIAVSGDKILSAMELDSYPLYAI 643
 | : || : : | | | | | | | | | | | | : : | | | |
 Db 582 LLLGIQYTDDDYYECSKGVWCRVGDFFPAIKSMGLNNLWI----DVFVMGVMLVGYRLMA- 636

Qy 644 YLIVIGLSGGFMVLYYVSLR 663
 : | | : | | |
 Db 637 -----YMALHRVKLR 646

RESULT 11

Q949Y4

ID Q949Y4 PRELIMINARY; PRT; 662 AA.

AC Q949Y4;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative ABC transporter protein.

GN F17M19.11.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

```

RP      SEQUENCE FROM N.A.
RA      Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA      Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA      Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA      Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA      Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
RA      Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA      Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA      Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT      "Full Length cDNA of gene F17M19.11 (GI:12324545).";
RL      Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AY050810; AAK92745.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR      GO; GO:0000166; F:nucleotide binding; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW      ATP-binding.
SO      SEQUENCE 662 AA; 72903 MW; CD5BC0853261BC45 CRC64;

```

Qy	44	PNTLEVRDLNYQVDLASQVPWFEEQLAQFKMPWTSPSCQNSCELGI-QNLSFKVRS----	97
		: : : : :: : :	
Db	37	PITLKFVDVCYRVKIHGM-----SNDSCNIKKLLGLKQKPSDETRSTEERT	82
Qy	98	-----GQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQGPSSPQLVRK	146
		: : : : : : : : ::	
Db	83	ILSGVTGMISPGEFMAVLGPGSGSGKSTLLNAVAGRLHGSNL-TGKILINDGKITKQTLKR	141
Qy	147	CVAHVRRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVG	206
		: : : : : :: : : : : :	
Db	142	-TGFVAQDDLLYPHLTVRETLVVFALLRLPRSLTRDVKIRAAESVISELGLTKCENTVVG	200
Qy	207	NMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSFTAHLNVLKTL SRLAKG-NRLV	265
		:: : : : : : : : : : : : : :	
Db	201	NTFIRGISGGERKRVSIAHELLINPSLLVLDEPTSGLDATAALRLVQTLAGLAHGKGKTV	260
Qy	266	LISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLT	325
		: : : : : : :: : : :: : :	
Db	261	VTSIHQPSSRVFQMFDTVLLLSEKCLFVGKGRDAMAYFESVGFSPAFFMNPADFLDLA	320
Qy	326	SIDRRSREQELATREK---AQSLAALFLEKVRDLDDFLWKAETKDLDDETCVESSVTPLD	382
		: : : : : : :	
Db	321	--NGVCQTDGVTEREKPNVRQTLVTAY-----DTLLAPQVK-----TCIEVSHFPQD	365
Qy	383	TNCLPSPTKMPGAVQQFTTLI-----RRQISNDFRDLPTLLIHGAEC	425
		:: : :	
Db	366	-NARFVKTRVNGG--GITTCIATWFSQLCILLHRLKERRHESFD-----LLRIFQVV	415

Qy	426	LMSMTIGFLSYFGHGSIQLSFMDTAALLFMI----	GALIPFNVILDVISKYSERAMLYYE	481
		: :: : : ::		
Db	416	AASILCGLMWW-HSDYR-DVHDLRGLLFFISIFWGVLP	SFNAVFTE----PQERAIFTRE	469
Qy	482	LEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTY	WLANLRPGLQPFLHFLLVVLVVF	541
		:: : : : : :		
Db	470	RASGMYTLSSYFMAHVLGSLSMELVLPASFLTFTY	WMVYLRPGIVPFLTLSVLLLYVLA	529
Qy	542	CRIMALAAAALLPTFHMASFFSNALYNSFYLAGG	FMINLSSLWTVPA---WISKVSFLRW	598
		: : : : : : : : :		
Db	530	SQGLGLALGAAIMDAKKASTIVTVTMLAFVLTG	GYVN-----KVPSGMVWMKYVSTTFY	584
Qy	599	CFEGLMKIQFSRRTYKMPLGNLTI	AVSGDKILSAMELDSYPLYA-----IYLIV	647
		: : : :: : :		
Db	585	CYRLLVAIQYG-----	SGEEILRMLGCDSKGKGASAATSAGCRFVEEEV	629
Qy	648	IGLSG-----	GFMVLYYVSLRFIK	666
		: ::		
Db	630	IGDVGMWTSVGVLELMFFGYRVLAYLALRRIK		661

084TH5

```

ID      Q84TH5                    PRELIMINARY;          PRT;    662 AA.
AC      Q84TH5;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Putative ABC transporter protein.
GN      AT1G71960.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA      Onodera C.S., Quach H.L., Tang C.C., Toriumi M., Wong C., Wu H.C.,
RA      Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
RA      Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
RA      Ecker J.R., Theologis A.;
RT      "Arabidopsis Open Reading Frame (ORF) Clones.";
RL      Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; BT005795; AAO64197.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR      GO; GO:0000166; F:nucleotide binding; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.

```

SQ SEQUENCE 662 AA; 72902 MW; AA84BD738D2D488A CRC64;

Query Match 20.2%; Score 708; DB 10; Length 662;

Best Local Similarity 31.2%; Pred. No. 2.2e-46;

Matches 216; Conservative 107; Mismatches 233; Indels 136; Gaps 23;

```
Qy      44 PNTLEVRDLNYQVDLASQVPWFELAQFKMPWTSPSCQNSCELGI-QNLSFKVRS----- 97
      | ||: |: |:| : :: || |:| | : ||
Db      37 PITLKFVDVCYRVKIHGM-----SNDSCNIKKLLGLKQKPSDETRSTEERT 82

Qy      98 -----GQMLAIIGSSGCGRASLLDVITGRGHGGKIKSGQIWINQPPSSPOLVRK 146
      |: :|:| | | |:|:|: : || || : :|:| | | : | :|:
Db      83 ILSGVTGMISPGEFMAVLGPGSGKSTLLNAVAGRLHGSNL-TGKILINDGKITKQTLKR 141

Qy     147 CVAHVRQHNQLLPNLTVRETIAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVG 206
      | | : | |:| || || |:| :| ||: : : : | ||:| | : | : | ||
Db     142 -TGFVAQDDLLYPHLTVRETIVFVALLRLPRSLTRDVKLRAAESVISELGLTKCENTVVG 200

Qy     207 NMYVRGLSGGERRRVSIGVQLLWNPILILDEPTSGLDSTAHNLVKTLSRLAKG-NRLV 265
      | :|:|:| || |:| | | :|:| || ||:| | | | : |
Db     201 NTFIRGISGGERKRVSIAGHELLINPSLLVLDEPTSGLDATAALRLVQTLAGLAHGKGTV 260

Qy     266 LISLHQPRSDIFRLFDLVLMTSGTPIYLGAQHMVQYFTAIGYPCPRYSNPADFYVDLT 325
      : |:|| | :|:| | ||:| : || : :| :|:| : || | | : ||
Db     261 VTSIHQPSSRVFQMFDTVLLLSEKCLFVGKGRDAMAYFESVGFSPAFAFPMNPADFLDLA 320

Qy     326 SIDRRSREQELATREK---AQSLAALFLEKVRDLDDFLWKAETKDLDCTVESSVTPLD 382
      : : : ||| |:| : | | : | ||:| | | |
Db     321 --NGVCQTDGVTEREKPNVRQTLVTAY-----DTLLAPQVK-----TCIEVSHFPQD 365

Qy     383 TNCLPSPTKMPGAVQQFTTLI-----RRQISNDFRDLPTLLIHGAEC 425
      | | :| | || | || | | | : :
Db     366 -NARFVKTRVNGG--GITTCIATWFSQLCILLHRLKERRHESFD-----LLRIFQVV 415

Qy     426 LMSMTIGFLYFGHGSIQLSFMDTAALLFMI---GALIPFNVILDVISKYSERAMLYYE 481
      |: | :|:| : | | || | | | | : ||:| |
Db     416 AASILCGLMWW-HSDYR-DVHDLRLGLLFFISIFWGVLPSPFNAVFTF---PQERAIFTRE 469

Qy     482 LEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLVWLTVFC 541
      |:| | | | :| | : ||:| |||: ||| : : | |
Db     470 RASGMYTLSSYFMAHVLGSLSMELVLPASFLTFTYWMVYLRPGIVPFLTLSVLLLYVLA 529

Qy     542 CRIMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSLWTVPA---WISKVSFLRW 598
      : : || | : || : | | ||: :| ||: | : | | :
Db     530 SQGLGLALGAAIMDAKASTIVTVMALAFVLTGGYYVN-----KVPSGMVWMKYVSTTFY 584

Qy     599 CFEGLMKIQFSRRTYKMPLGNLTIAVSGDKILSAMELDSYPLYA-----IYLIV 647
      |: |: ||: ||:| | : || : |
Db     585 CYRLLVAIQYG-----SGEEILRMLGCDSKKGQGASAATSAGCRFVEEEV 629

Qy     648 IGLSG-----GFMVLYYVSLRFIK 666
      || | |: || |:| | |
Db     630 IGDVGMWTSVGVLFMLFFGYRVLAYLALRRIK 661
```

RESULT 13

Q9C8W6

ID Q9C8W6 PRELIMINARY; PRT; 609 AA.
AC Q9C8W6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative ABC transporter.
GN F17M19.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
DR EMBL; AC021665; AAG52231.1; -.
DR PIR; E96742; E96742.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding.
SQ SEQUENCE 609 AA; 67007 MW; 65D11A874E5C0B61 CRC64;

Query Match 20.0%; Score 700; DB 10; Length 609;
Best Local Similarity 31.8%; Pred. No. 8.2e-46;
Matches 210; Conservative 103; Mismatches 225; Indels 122; Gaps 22;

Qy 76 TSPSCQNSCELGI-QNLSFKVRS-----GQMLAIIGSSGCGRASLLDVI 118
:: || ||: | | : || |: :||:| || |::||: :

RC STRAIN=Ax4;
 RA Anjard C., Loomis W.F.;
 RT "Evolution of the ABC transporters of Dictyostelium.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AF482380; AAL91485.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 801 AA; 90052 MW; CCC4F0036CB195A3 CRC64;

Query Match 19.8%; Score 695.5; DB 5; Length 801;
 Best Local Similarity 27.9%; Pred. No. 2.7e-45;
 Matches 187; Conservative 134; Mismatches 230; Indels 119; Gaps 20;

Qy	88	IQNLSFKVRSQMLAIIGSSGCGRASLLDVITGRGHGKIK-SGQIWINGQPSSPQLVRK	146
		: :: : : : : : :: : :: : :	
Db	139	LTNINGHIESGTIFAIMGPSGAGKTLLDIL---AHRLNINGSGMTYLNKNSDFNIFKK	195
Qy	147	CVAHVRQHNQLLPNLTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVG	206
		: : : : :: : : : : : :	
Db	196	LCGYVTQSDSLMPSLTVRETLNFYAQLKMPRDVPLKEKLQRVQDIIDEMGLNRCADTLVG	255
Qy	207	--NMYVRGLSGGERRRVSIGVQLLWNPGLILILDEPTSGLDSTAHNLVKTLSRLAKGNRL	264
		: : : : : : : : : : :	
Db	256	TADNKIRGISGGERRRVTISIELLTGPSVILLDEPTSGLDASTSFYVMSALKKLAKSGRT	315
Qy	265	VLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDL	324
		:: : : : : : : : : :	
Db	316	IICTIHQPRSNIDYDMFDNLLLLGDGNTIYYGKANKALEYFNANGYHCSEKTNPADFFDL	375
Qy	325	-----TSID-----	328
		:::	
Db	376	INTQVEDQADSDDDDYNDEEEEIGGGGGSGGGAGGIEDIGISISPTMNGSAVDNIKNNE	435
Qy	329	-RRSREQELATREKAQSLAALFLEKVRDLDDFLWKAETKDLDDETCVESSVTPLD----T	383
		:: : : : : : : : : : : :	
Db	436	LKQQQQQQQQQQSTDGRRARRIKKLTKEEMVILKKEYPNSEQGLRVNETLDNISKENRT	495
Qy	384	NCLPSPTKMPGAVQQFTTLIRQISNDFRDLPTLLIHGAACLMSMTIGFLYF----GHG	439
		: : : : : : : : : : :	
Db	496	DFKYEKTRGPNFLTQFSLLLGREVTNAKRHPMAFKVNLIQAIFQGLLCGIVYYQLGLGQS	555
Qy	440	SIQLSFMDTAALLFMI-GALIP-----FNVILDVISKCYSERAMLYYELEDGLYTTGPYF	493
		: : : : : : : :	
Db	556	SVQ---SRTGVVAFIIMGVSFPAVMSTIHVFPDVITIFLKDRA-----SGVYDTLPFF	605

Qy	494	FAK-----ILGELPEHCAYIIYGMPTYWLANLRP----GLQPFLLHFLLVWLVVFCR	543
Db	606	LAKSFMDACIAVLLPMVTATIV-----YWMTNQVRDPFYSAAPFFRFVLMVLASQTCL	659
Qy	544	IMALAAAALLPTFHMASFFSNALYNSFYLAGGFMINLSSL--WTVPAWISKVSFLRWCFE	601
		: : : : : : : : : : : : :	
Db	660	SLGVLISSSVPNVQVGTAVAPLIVILFFLFSGGFINLNDVPGWLIV--WFPYISFFRYMIE	717
Qy	602	GLMKIQFS----RRTYKMPLGNLTIAVSGDKILSAM--ELDSY--PLYAIYLIVIGLSGG	653
		: : : : : : : : : : :	
Db	718	AAVINAFKDVHFTCTDSQKIGGVCVPQYGNNVNIENMGYDIDHFWRNVWILVLYII---G	773
Qy	654	FMVLYYVSLR	663
		: :	
Db	774	FRVLTFLLVK	783

07TSR8

AC 07TSR8;

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DE ATP-binding cassette sub-family G member 5.

OS Mus musculus (Mouse).

OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.

CH	NSD
RN	[1]

RC STRAIN=I/LnJ: TISS

BA Paigen B.:

RT Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbr

RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.

KW ATP-binding.

Query Match 19.6%; Score 686.5; DB 11; Length 652;

Matches 188; Conservative 128; Mismatches 242; Indels 99; Gaps 16;

QY 45 NTLEVRDLNYQVDLASQV-PWFEQLAQFKMPWTSPSCQNSELGI-QNLSFKVRSGQMLA 102
 ::|||::||:::||||| ||||:|::|:||||:

QY 103 IIGSSGCRASLLDVITGRGHGGKIKSGQIWINGQPSSPQLVRKCAVHVRQHNQLLPNLT 162
|:|||| |: :|| |:| |::|| :| :|| |:|

Qy 163 VRETLAFIAQMRLPRTFSQAQRDKRVEDVIAELRLRQCADTRVGNNMYVRGLSGGERRRVSV 222
 ||||| : | : | : | : || | | | | | : : : | : | |||||

Db 145 VRETLRYTAMLALCRS-SADFYNKKVEAVMTELSLSHVADQVIGSYNFGGISSGERRRVS 203
 Qy 223 IGVQLLWNPGLILDEPTSGLDSTAHNLVKTL SRLAKGNRLVLISLHQPRSDIFRLFDL 282
 | ||| :| ::||| ||| ||: :| |: ||: |:|::: |||::: ||
 Db 204 IAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELARRDRIVIVTIHQPRSELFQHFDDK 263
 Qy 283 VLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRYSNPADFYVDLTSIDRRSREQELATREKA 342
 : ::| | : : | : | : ||| | ||| : ||| : | : ||| : | : :
 Db 264 IAILTYGELVFCGTPEEMLGFFNCGYPCPEHSNPFDFYMDLTSVDTQSREREIETYKRV 323
 Qy 343 QSLAALFLEKVRDLDDFLWKAETKDLDEDTCESSVTPLDTNCLPSPTK-MPGAVQQFTT 401
 | | : | | | : : : : : : | : | | | :
 Db 324 QMLESAPKE-----SDIYHKI-LENIERARYLKTLP-----VPFKTKDPPGMFGKLG 371
 Qy 402 LIRRQISNDFRDLPTLLIHGAEACLSMTIGF--LYFGHGSIQLSFMDTAALLFMIGALI 459
 |: || | | : : : : : | : : | | : : : : | ||:
 Db 372 LLRRVTRNLMRNKQAVIMRLVQNLMGLFLIFYLLRVQNNLTGKAVQDRVGLLYQFVGAT 431
 Qy 460 PFNVILDVISKCYSERAMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLA 519
 |: :|: : : ||: | : ||| | : | || :|: ||
 Db 432 PYTGMLNAVNLFPMMLRAVSDQESQDGLYHKWQMLLAYVLHALPFSIIATVIFSSVCYWT 491
 Qy 520 NLRPGLQPFLHFLLVVLVVFCCRIMALAAAALLPTFHMAFFSNAL----- 566
 | | : | : ||| : | : |
 Db 492 GLYPEVARF-----GYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSI 534
 Qy 567 -----YNSFYLAGGFMINLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLT 621
 : : ||: |: : : :| :| | |: :| | : | |
 Db 535 VALLSISGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVNEF-----YGL---NFT 587
 Qy 622 IAVSGDKILSAMELDSYPLYAI-----YLIVIGLSGGFMVL 657
 | : | :|: || :||: | :|
 Db 588 CGESNTTML-----NHPMCAITQGVFIEKTCPGATSRFTANFLILYGFIPALVIL 638

Search completed: February 27, 2004, 07:15:30
 Job time : 39.3606 secs